Quantitative Stability Analysis for Distributionally Robust Optimization With Moment Constraints*

Jie Zhang† Huifu Xu‡ and Li-wei Zhang§

September 17, 2015

Abstract. In this paper we consider a broad class of distributionally robust optimization (DRO for short) problems where the probability of the underlying random variables depends on the decision variables and the ambiguity set is defined through parametric moment conditions with generic cone constraints. Under some moderate conditions including Slater type conditions of cone constrained moment system and Hölder continuity of the underlying random functions in the objective and moment conditions, we show local Hölder continuity of the optimal value function of the inner maximization problem w.r.t. the decision vector and other parameters in moment conditions, local Hölder continuity of the optimal value of the whole minimax DRO w.r.t the parameter. Moreover, under the second order growth condition of the Lagrange dual of the inner maximization problem, we demonstrate and quantify the outer semicontinuity of the set of optimal solutions of the minimax DRO w.r.t variation of the parameter. Finally we apply the established stability results to two particular class of DRO problems.

Key words. Distributionally robust optimization, moment conditions with cone constraints, Hölder continuity of the optimal value function, outer semicontinuity of the set of optimal solutions, quantitative stability analysis

1 Introduction

Consider the following distributionally robust optimization problem:

$$(P_u) \qquad \begin{array}{ll} \min\limits_{x} & \sup\limits_{P\in\mathcal{P}(x,u)} \mathbb{E}_P[f(x,\xi(\omega))] \\ \text{s.t.} & x\in X, \end{array} \tag{1.1}$$

where X is a closed set of $\mathbb{R}^n$ and U is a Banach space, $f:\mathbb{R}^n\times\mathbb{R}^k\to\mathbb{R}$ is a continuous function, $\xi:\Omega\to\Xi$ is a vector of random variables defined on probability space $(\Omega,\mathcal{F},P)$ with support set $\Xi\subset\mathbb{R}^k$, $\mathcal{P}(x,u)$ is a set of distributions which contains the true probability distribution of random variable ξ, and $\mathbb{E}_P[\cdot]$ denotes mathematical expectation with respect to probability measure $P\in\mathcal{P}$.

One of the key ingredients in this formulation is the set of probability distributions $\mathcal{P}(x,u)$. For each fixed $(x,u)\in X\times U$, we confine our discussions in this paper to the case that $\mathcal{P}(x,u)$ is constructed through moment conditions:

$$\mathcal{P}(x,u) := \{P\in\mathscr{P} : \mathbb{E}_P[\Psi(x,u,\xi(\omega))]\in\mathcal{K}\}, \tag{1.2}$$

*The research is supported by EPSRC grant EP/M003191/1, NSFC Grant #11201210 and CPSF Grant #2014M560200.

†School of Mathematics, Liaoning Normal University, Dalian 116029, China. (zhangjie04212001@163.com).

‡School of Mathematics, University of Southampton, Southampton, SO17 1BJ, UK. (h.xu@soton.ac.uk), Haitian Scholar, Dalian University of Technology.

§Institute of ORCT, School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, China. (lwzhang@dlut.edu.cn).

where Ψ is a random mapping consisting of vectors and/or matrices with measurable random components, the mathematical expectation of Ψ is taken w.r.t. each component of Φ, $\mathscr{P}$ denotes the set of all probability distributions/measures in the space space $(\Omega, \mathcal{F}, P)$, and $\mathcal{K}$ is a closed convex cone in the Cartesian product of some finite dimensional vector and/or matrix spaces. Throughout the paper, we assume without loss of generality that $\mathcal{P}(x,u) \neq \emptyset$ for each $x \in X$ and $u \in U$.

Note that if we consider $(\Xi, \mathscr{B})$ as a measurable space equipped with Borel sigma algebra $\mathscr{B}$, then $\mathcal{P}(x,u)$ may be viewed as a set of probability measures defined on $(\Xi, \mathscr{B})$ induced by the random variate ξ. Following the terminology in the literature of robust optimization, we call $\mathcal{P}$ the *ambiguity set* which indicates ambiguity of the true probability distribution of ξ in this setting. To ease notation, we will use ξ to denote either the random vector $\xi(\omega)$ or an element of $\mathbb{R}^k$ depending on the context.

Obviously the ambiguity set depends on Ψ. Two special cases which might be of interest: (a) Ψ is a vector valued function, (1.2) collapses to classical moment problem. In that case, u often represents moments such as the mean and variance. (b) Ψ is a matrix, (1.2) reduces to matrix moments and the parameter u may represent the convariance matrix of some random vectors. The research on DRO with classical moments have been well documented, see for instance [5, 19] and the references therein. In all these works, the ambiguity set is independent of decision vector.

Our interest here is in the case when Ψ takes a general form that depends on both x and u. We investigate the impact of variation of x and u on problem (P_u). The dependence on x makes the DRO problem significantly more complicated mathematically but it enables us to cover a wider range of applications than the existing DRO models. This is even so when $\mathcal{P}(x,u)$ reduces to a singleton $\{P(x.u)\}$, the DRO problem collapses to a one stage stochastic program

$$\begin{array}{ll} \min\limits_{x} & \mathbb{E}_{P(x,u)}[f(x,\xi(\omega))] \\ \text{s.t.} & x \in X, \end{array} \tag{1.3}$$

which differs from standard one stage SP models for the dependence of $P(x,u)$ on x. The new modelling paradigm means that any decision will have a direct impact on the likelihood of the underlying random events that occur after the decision is taken. This is the case in many engineering decision making problems where the likelihood of failure of a project is closely related to structural design. Similar examples can also be found in management sciences where a decision maker's efforts may affect the likelihood of success or failure of his objective. Indeed, in his influential seminal work on principal agent model, Mirrlees [12] explicitly considers dependence of the probability distribution of the agent's success on his/her efforts. The parameter u may be used to describe the parameters which characterize a probability distribution such as the mean and/or variance. Therefore u often takes a functional form of x. We will give more details in Section 4.

The DRO model (1.1) can also be motivated from robust formulation of one stage stochastic program with expected stochastic constraints

$$\begin{array}{ll} \min\limits_{x} & \mathbb{E}_P[f(x,\xi(\omega))] \\ \text{s.t.} & x \in X, \\ & \mathbb{E}_P[g(x,\xi)] \leq 0, \end{array} \tag{1.4}$$

where $g : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^m$ is a vector-valued function. If we don't know the true probability distribution P_0, but it is possible to construct a set of distributions $\mathcal{P}_0(u)$ such that $P_0 \in \mathcal{P}_0(u)$, then it might be sensible to consider the following distributionally robust formulation of problem (1.4)

$$\begin{array}{lll} \min\limits_{x \in X} & \sup\limits_{P \in \mathcal{P}_0(u)} & \mathbb{E}_P[f(x,\xi(\omega))] \\ & \text{s.t.} & \mathbb{E}_P[g(x,\xi)] \leq 0. \end{array} \tag{1.5}$$

The parameter u in $\mathcal{P}_0(u)$ may be interpreted as moments (e.g., the mean and variance of ξ or some reference random variables) if $\mathcal{P}_0(u)$ is specified via moment conditions.

Let

$$\mathcal{P}(x,u) := \{P \in \mathcal{P}_0(u) : \mathbb{E}_P[g(x,\xi)] \le 0\}.$$

Then we can effectively reformulate problem (1.5) into DRO problem (1.1). It is easy to verify that problem (1.5) is indeed equivalent to

$$\begin{array}{ll} \min\limits_{x \in X} & \sup\limits_{P \in \mathcal{P}_0(u)} \mathbb{E}_P[f(x,\xi(\omega))] \\ \text{s.t.} & \sup\limits_{P \in \mathcal{P}_0(u)} \mathbb{E}_P[g(x,\xi)] \le 0, \end{array} \tag{1.6}$$

which is widely known in the literature of robust optimization.

The purpose of this paper is to investigate how a small variation of the decision vector x and parameter u affects the optimal value and the optimal solution of the inner maximization problem of (P_u). This kind of analysis may be traced down to the earlier research by Breton and Hachem [3, 4] and Takriti and Ahmed [22] whose stability analysis are carried out for some distributionally robust optimization problems with finite discrete probability distributions. Riis and Andersen [16] extend such analysis to continuous probability distributions. More recently, Sun and Xu [21] present a comprehensive asymptotic analysis of DRO w.r.t variation of the underlying ambiguity set under total variation metric and pseudo metric. In all these works, the dependence of the ambiguity set is confined to parameter u. Moreover, the analysis is qualitative rather than quantitative and it is carried out directly for the minimax DRO.

In this paper, we advance the research by considering dependence of the ambiguity set on the decision vector x as motivated in the earlier discussions and presenting a detailed *quantitative* stability analysis for (P_u). Moreover, differing from the existing research in the literature, part of our stability analysis (for the optimal solutions) is carried out through the Lagrange dual of the inner maximization problem, a popular formulation in DRO for its numerical solution.

The rest of the paper are organized as follows. We start in Section 2 with quantitative stability analysis of the ambiguity set $\mathcal{P}(x,u)$ w.r.t. variation of the decision vector x and parameter u. This is essentially about deriving Hölder continuity of $\mathcal{P}(x,u)$ as a set-valued mapping under total variation metric (Theorem 2.1). A key step towards this is to establish Hoffman's lemma for the cone constrained moment system (Lemma 2.1). Section 3 presents a detailed stability analysis for the optimal value of the inner maximization problem as a function of (x,u) and the optimal value of (P_u) as a function of u. Specifically, under some moderate conditions such as Slater type conditions of the cone constrained moment system (1.2) and Hölder continuity of the underlying random functions in the objective and the moment system, we show local Hölder continuity of the optimal value of the inner maximization problem as a function of (x,u) (Theorem 3.1), local Hölder continuity of the optimal value of (P_u) w.r.t u (Theorem 3.2). Section 4 discusses stability of the optimal solutions. Under the second order growth condition of the Lagrange dual of the inner maximization problem of (P_u), we demonstrate and quantify outer semicontinuity of the optimal solution set mapping to both (P_u) and the Lagrange dual of its inner maximization problem. Finally we apply the established stability results to two particular classes of DRO problems in Section 5.

Throughout the paper, we us the following notation. By convention, we use $\mathbb{R}^{n\times n}$ and $S^{n\times n}$ to denote the space of all $n\times n$ matrices and symmetric matrices, and $S_+^{n\times n}$ and $S_-^{n\times n}$ the cone of positive semi-definite and negative semi-definite symmetric matrices. For matrices $A, B \in \mathbb{R}^{n\times n}$, we write $A \bullet B$ for the Frobenius inner product, that is $A \bullet B := \mathrm{tr}(A^T B)$, where "tr" denotes the trace of a matrix and the superscript T denotes transpose. Moreover, we use standard notation $\|A\|_F$ for the Frobenius norm of A, that is, $\|A\|_F := (A \bullet A)^{1/2}$, $\|x\|$ for the Euclidean norm of a vector x in $\mathbb{R}^n$, $\|x\|_\infty$ for the infinity norm and $\|\psi\|_\infty$ for the maximum norm of a real valued measure function $\psi : \Xi \to \mathbb{R}$. For a Banach space X, we write $B(x, \delta)$ for the closed ball with center $x \in X$ and radius δ and $\mathcal{B}$ for the closed unit ball in X. For a set $S \subseteq X$, $\mathrm{int}\, S$ denotes the interior S, $d(x, S) := \inf_{x' \in S} \|x - x'\|$ denotes the distance from point $x \in X$ to a set $S \subset X$. For two sets $S_1, S_2 \subset X$,

$$\mathbb{D}(S_1, S_2) := \inf\{t \geq 0 : S_1 \subset S_2 + t\,\mathcal{B}\}$$

signifies the deviation of S_1 from S_2 and $\mathbb{H}(S_1, S_2) := \max(\mathbb{D}(S_1, S_2), \mathbb{D}(S_2, S_1))$ denotes the Hausdorff distance between the two sets.

2 Stability of the ambiguity set

In order to study stability of the DRO problem (P_u), we need to investigate some topological properties of the optimal value of its inner maximization problem. By taking into account of the structure of the ambiguity set in (1.2), we write the inner maximization problem as

$$(P_{x,u}) \qquad \begin{array}{cl} \sup\limits_{P \in \mathscr{P}} & \mathbb{E}_P[f(x, \xi)] \\ \text{s.t.} & \mathbb{E}_P[\Psi(x, u, \xi)] \in \mathcal{K}. \end{array} \tag{2.7}$$

Note that in this formulation, the probability measure P is a decision variable and the set of feasible solutions is the ambiguity set $\mathcal{P}(x, u)$. This is an infinite dimensional parametric program with parameters (x, u).

Analogous to standard stability analysis in parametric programming, we begin by looking into variation of the feasible set of the maximization problem (2.7) w.r.t. change of the parameters, i.e., the continuity of $\mathcal{P}(\cdot, \cdot)$ as a set-valued mapping from $X \times U$ to $\mathscr{P}$. To this end, we introduce an appropriate metric which can be used to quantify the change of probability measures in the space $\mathscr{P}$.

2.1 Total variation metric

In probability theory, various metrics have been introduced to quantify the distance/difference between two probability measures; see [1, 10]. Here we adopt total variation metric which subsumes Lipschitz metric and some other metrics; see [10] and references therein.

Definition 2.1 Let $P, Q \in \mathscr{P}$ and $\mathscr{H}$ denote the set of measurable functions defined in the probability space $(\Xi, \mathscr{B})$. The *total variation metric* between P and Q is defined as (see e.g., page 270 in [1])

$$d_{TV}(P, Q) := \sup_{h \in \mathscr{H}} \left| \mathbb{E}_P[h(\xi)] - \mathbb{E}_Q[h(\xi)] \right|, \tag{2.8}$$

where

$$\mathscr{H} := \left\{ h : \mathbb{R}^k \to \mathbb{R} | h \text{ is } \mathscr{B} \text{ measurable}, \sup_{\xi \in \Xi} |h(\xi)| \leq 1 \right\}, \tag{2.9}$$

and *total variation norm* as

$$||P||_{TV} = \sup_{\|\phi\|_\infty \leq 1} \left| \mathbb{E}_P[\phi(\xi)] \right|.$$

If we restrict the measurable functions in set $\mathscr{H}$ to be uniformly Lipschitz continuous, that is,

$$\mathscr{H} = \left\{ h : \sup_{\xi \in \Xi} |h(\xi)| \leq 1,\ L_1(h) \leq 1 \right\}, \tag{2.10}$$

where

$$L_1(h) = \inf\{L : |h(\xi') - h(\xi'')| \leq L\|\xi' - \xi''\|, \forall \xi', \xi'' \in \Xi\},$$

then $d_{TV}(P, Q)$ is known as bounded Lipschitz metric, see e.g. [14] for details. In the case when $L_1(h) \leq 1$, $d_{TV}(P, Q)$ reduces to Wasserstein (or Kantorovich) metric.

Using the total variation norm, we can define the distance from a point to a set, deviation from one set to another and Hausdorff distance between two sets in the space of $\mathscr{P}$. Specifically, let

$$d_{TV}(Q, \mathcal{P}) := \inf_{P \in \mathcal{P}} d_{TV}(Q, P),$$

$$\mathbb{D}_{TV}(\mathcal{P}', \mathcal{P}) := \sup_{Q \in \mathcal{P}'} d_{TV}(Q, \mathcal{P})$$

and

$$\mathbb{H}_{TV}(\mathcal{P}', \mathcal{P}) := \max\{\mathbb{D}_{TV}(\mathcal{P}', \mathcal{P}), \mathbb{D}_{TV}(\mathcal{P}, \mathcal{P}')\}.$$

Here $\mathbb{H}_{TV}(\mathcal{P}', \mathcal{P})$ defines Hausdorff distance between $\mathcal{P}'$ and $\mathcal{P}$ under the total variation metric in space $\mathscr{P}$. It is easy to observe that $\mathbb{H}_{TV}(\mathcal{P}', \mathcal{P}) = 0$ implies $\mathbb{D}_{TV}(\mathcal{P}', \mathcal{P}) = 0$ and

$$\inf_{Q \in \mathcal{P}} \sup_{h \in \mathscr{H}} |\mathbb{E}_P[h(\xi)] - \mathbb{E}_Q[h(\xi)]| \to 0$$

for any $P \in \mathcal{P}$.

2.2 Hoffman's lemma

We now return to stability analysis of ambiguity set $\mathcal{P}(x, u)$, i.e., change of $\mathcal{P}(x, u)$ under the total variation metric w.r.t. as (x, u) varies. Let $\mathscr{M}_+$ denote the positive linear space of all signed measures generated by $\mathscr{P}$, let

$$\langle P, \Psi(x, u) \rangle := \int_\Xi \Psi(x, u) P(d\xi).$$

Problem (2.7) can be equivalently formulated as

$$(P_{x,u}) \quad \begin{array}{ll} \sup\limits_{P \in \mathscr{M}_+} & \langle P, f(x, \xi) \rangle \\ \text{s.t.} & \langle P, \Psi(x, u, \xi) \rangle \in \mathcal{K}, \\ & \langle P, \Psi(x, u, \xi) \rangle = 1. \end{array} \tag{2.11}$$

An advantage of the reformulation is that it enables us to see more clearly that the inner maximization problem is indeed a *linear* parametric program w.r.t P. We will exploit both formulations interchangeably in later discussions depending on the context.

The first technical result to be established is characterization of the distance from a point to the set of feasible solutions of (2.7), i.e., the distance of any probability measure $Q \in \mathscr{P}$ to set $\mathcal{P}(x, u)$ under the total variation metric in terms of the residual of the system (1.2). The result is a generalization of the well known Hoffman's lemma for linear systems of inequalities in finite dimensional space ([20, Theorem 7.11].

Lemma 2.1 (Hoffman's lemma for the moment problem (1.2)) *Let (x_0, u_0) be fixed. Assume that*

$$0 \in \text{int}\,\{(\langle P, \Psi(x_0, u_0, \xi)\rangle - \mathcal{K}) : P \in \mathscr{P}\} \tag{2.12}$$

and Ψ is uniformly continuous w.r.t. (x, u) near (x_0, u_0). Here and later on "int" denotes the interior of a set. Then there exists a positive constant C such that

$$d_{TV}(Q, \mathcal{P}(x, u)) \leq Cd(\mathbb{E}_Q[\Psi(x, u, \xi)], \mathcal{K}) \tag{2.13}$$

for any $Q \in \mathscr{P}$ and $(x, u) \in X \times U$ close to (x_0, u_0). For brevity, here and later on, we write 0 for a tuple with 0 components corresponding to those of Ψ.

Proof. The proof is similar to Hoffman's lemma for classical equality and inequality constrained moment problem (see [21, Lemma 4.1]) and the thrust of the proof is to utilize Shapiro's duality theorem [19, Proposition 3.4]. Here we give the details of the proof for completeness. We proceed in two steps.

Step 1. We show that there exists a positive constant C_0 such that

$$d_{TV}(Q, \mathcal{P}(x_0, u_0)) \leq C_0 d(\mathbb{E}_Q[\Psi(x_0, u_0, \xi)], \mathcal{K}) \tag{2.14}$$

for any $Q \in \mathscr{P}$. Let $P \in \mathcal{P}(x_0, u_0)$. Since $\Psi(x_0, u_0, \xi)$ is continuous in ξ, each component of $\Psi(x_0, u_0, \xi)$ is P-integrable function. By the definition of the total variation norm, $||P||_{TV} = \sup_{\|\phi\|_\infty \leq 1}\langle P, \phi(\xi)\rangle$. Moreover, for any fixed $Q \in \mathscr{P}$, by the definition of the total variation metric

$$\begin{aligned} d_{TV}(Q, \mathcal{P}(x_0, u_0)) &= \inf_{P \in \mathcal{P}(x_0, u_0)} d_{TV}(Q, P) \\ &= \inf_{P \in \{P : \mathbb{E}_P[\Psi(x_0, u_0, \xi)] \in \mathcal{K}\}} \sup_{\|\phi\|_\infty \leq 1} \langle Q - P, \phi(\xi)\rangle \\ &= \sup_{\|\phi\|_\infty \leq 1} \inf_{P \in \{P : \mathbb{E}_P[\Psi(x_0, u_0, \xi)] \in \mathcal{K}\}} \langle Q - P, \phi(\xi)\rangle, \end{aligned}$$

where the exchange of infimum and supremum is justified by [9, Theorem 1]. In what follows, we derive the Lagrange dual of $\inf_{P \in \{P : \mathbb{E}_P[\Psi(x_0, u_0, \xi)] \in \mathcal{K}\}} \langle Q - P, \phi(\xi)\rangle$. Let

$$\mathcal{L}(P, \Lambda) := \langle Q - P, \phi(\xi)\rangle - \Lambda \bullet (\mathbb{E}_P[\Psi(x_0, u_0, \xi)])$$

and

$$\mathcal{K}^* := \{M : \langle M, V\rangle \geq 0, \forall\ V \ \in \mathcal{K}\}, \tag{2.15}$$

where $A \bullet B$ denotes Frobenius product of matrices A and B. Then

$$\mathcal{L}(P, \Lambda) := \langle Q - P, \phi(\xi) - \Lambda \bullet \Psi(x_0, u_0, \xi)\rangle - \Lambda \bullet \langle Q, \Psi(x_0, u_0, \xi)\rangle.$$

Since $Q - P$ is a signed measure, $\phi(\xi) - \Lambda \bullet \Psi(x_0, u_0, \xi) = 0$ almost surely. Consequently, the Lagrange dual can be written as

$$\sup_{\Lambda \in \mathcal{K}^*, \phi(\xi) = \Lambda \bullet \Psi(x_0, u_0, \xi)} -\Lambda \bullet (\mathbb{E}_Q[\Psi(x_0, u_0, \xi)]).$$

On the other hand, under condition (2.12), it follows by [19, Proposition 3.4] that

$$\inf_{P \in \{P : \mathbb{E}_P[\Psi(x_0, u_0, \xi)] \in \mathcal{K}\}} \langle Q - P, \phi(\xi)\rangle = \sup_{\Lambda \in \mathcal{K}^*, \phi(\xi) = \Lambda \bullet \Psi(x_0, u_0, \xi)} -\Lambda \bullet (\mathbb{E}_Q[\Psi(x_0, u_0, \xi)]) \tag{2.16}$$

and hence

$$d_{TV}(Q,\mathcal{P}) = \sup_{\Lambda\in\mathcal{K}^*, \|\Lambda\bullet\Psi(x_0,u_0,\xi)\|_\infty\leq 1} -\Lambda\bullet(\mathbb{E}_Q[\Psi(x_0,u_0,\xi)]). \tag{2.17}$$

Let $\mathcal{S}^*$ denote the set of optimal solutions to the program

$$\sup_{\Lambda\in\mathcal{K}^*, \|\Lambda\bullet\Psi(x_0,u_0,\xi)\|_\infty\leq 1} -\Lambda\bullet(\mathbb{E}_Q[\Psi(x_0,u_0,\xi)]). \tag{2.18}$$

In what follows, we show that the intersection of $\mathcal{S}^*$ with some bounded set (to be defined shortly) is nonempty for any Q. In other words the program always has an optimal solution in a specified bounded set. Let $\mathcal{S}$ denote the feasible set of program (2.18), that is,

$$\mathcal{S} := \{\Lambda\in\mathcal{K}^* : \|\Lambda\bullet\Psi(x_0,u_0,\xi)\|_\infty\leq 1\}.$$

Since $\mathcal{S}$ is a set in $\mathcal{K}^*$, it can be written as a Minkowski sum $\mathcal{S}_0+\mathcal{C}$, where $\mathcal{S}_0$ is a bounded convex set and $\mathcal{C}$ is a convex cone. We want to show

$$\mathcal{S}_0\bigcap\mathcal{S}^*\neq\emptyset \tag{2.19}$$

for any $Q\in\mathcal{P}$. By definition, for any $\Lambda\in\mathcal{S}$, there exist $\Lambda_s\in\mathcal{S}_0$ and $\Lambda_c\in\mathcal{C}$ such that $\Lambda=\Lambda_s+\Lambda_c$. Obviously for any positive number M, $M\Lambda_c\in\mathcal{C}$ and hence $\Lambda_s+M\Lambda_c\in\mathcal{S}$. Moreover

$$\Lambda_c\bullet\Psi(x_0,u_0,\xi)=0. \tag{2.20}$$

To see this, if there exist $\Lambda_c^0\in\mathcal{C}$ and $\xi_0\in\Xi$ such that $\Lambda_c^0\bullet\Psi(x_0,u_0,\xi_0)\neq 0$. Then for any Λ_s, we can find a positive number M sufficiently large such that

$$\|(\Lambda_s+M\Lambda_c^0)\bullet\Psi(x_0,u_0,\xi)\|_\infty \geq |(\Lambda_s+M\Lambda_c^0)\bullet\Psi(x_0,u_0,\xi_0)|>1,$$

which means $\Lambda_s+M\Lambda_c^0\notin\mathcal{S}$, a contradiction to our argument that we have just shown. By (2.20), we arrive at

$$\Lambda_c\bullet\mathbb{E}_Q[\Psi(x_0,u_0,\xi)]=0 \tag{2.21}$$

for any $\Lambda_c\in\mathcal{C}$. Let Λ^* denote the optimal solution of program (2.18). Let $\Lambda^*\in\mathcal{S}_0\bigcap\mathcal{S}^*$ and

$$C:=\max_{\Lambda\in\mathcal{S}_0}\|\Lambda\| \tag{2.22}$$

Observe that $\mathcal{S}_0$ is only determined by system $\|\Psi(x_0,u_0,\xi)\bullet\Lambda\|_\infty\leq 1$. Thus, C is independent of $\mathbb{E}_Q$. Moreover,

$$\begin{aligned}
\sup_{\Lambda\in\mathcal{K}^*, \|\Lambda\bullet\Psi(x_0,u_0,\xi)\|_\infty\leq 1} -\Lambda\bullet(\mathbb{E}_Q[\Psi(x_0,u_0,\xi)]) &= -\Lambda^*\bullet(\mathbb{E}_Q[\Psi(x_0,u_0,\xi)])\\
&= \|\Lambda^*\|(-\Lambda^*/\|\Lambda^*\|)\bullet(\mathbb{E}_Q[\Psi(x_0,u_0,\xi)])\\
&\leq \|\Lambda^*\|d(\mathbb{E}_Q[\Psi(x_0,u_0,\xi)],\mathcal{K})\\
&\leq \max_{\Lambda\in\mathcal{S}_0}\|\Lambda\|d(\mathbb{E}_Q[\Psi(x_0,u_0,\xi)],\mathcal{K})\\
&\leq Cd(\mathbb{E}_Q[\Psi(x_0,u_0,\xi)],\mathcal{K}).
\end{aligned}$$

In the first inequality, we are using the following fact: if $-\Lambda^*/\|\Lambda^*\|$ and $\mathbb{E}_Q[\Psi(x_0,u_0,\xi)]$ are in an obtuse angle, then the product is negative and if they are in a sharp angle, then

$$(-\Lambda^*/\|\Lambda^*\|)\bullet(\mathbb{E}_Q[\Psi(x_0,u_0,\xi)])\leq d(\mathbb{E}_Q[\Psi(x_0,u_0,\xi)],\mathcal{K}).$$

Combining the discussions above, we have

$$d_{TV}(Q,\mathcal{P}(x_0,u_0)) = -\Lambda^* \bullet (\mathbb{E}_Q[\Psi(x_0,u_0,\xi)]) \le Cd(\mathbb{E}_Q[\Psi(x_0,u_0,\xi)],\mathcal{K}).$$

Step 2. From (2.12), we can easily show that by uniform continuity of Ψ w.r.t. x, u,

$$0 \in \text{int}\,\{(\langle P,\Psi(x,u,\xi)\rangle - \mathcal{K}) : P \in \mathscr{P}\} \tag{2.23}$$

for (x,u) close to (x_0,u_0). Then for fixed (x,u) close to (x_0,u_0), similarly to the proof in step 1, there exists $\widehat{C}$ depending on (x,u) such that

$$d_{TV}(Q,\mathcal{P}(x,u)) \le \widehat{C}d(\mathbb{E}_Q[\Psi(x,u,\xi)],\mathcal{K}) \tag{2.24}$$

for any $Q \in \mathscr{P}$. To complete the proof, we need to show that $\widehat{C}$ is uniformly bounded for all (x,u) close to (x_0,u_0). Let

$$\mathcal{F}(x,u) := \{\Lambda \in \mathcal{K}^* : \|\Psi(x,u,\xi)\bullet\Lambda\|_\infty \le 1\}.$$

From the proof of Step 1, it is enough to show that $\mathcal{F}(x,u)$ is uniformly bounded in a neighborhood of (x_0,u_0). Assume for the sake of a contradiction that $\mathcal{F}(x,u)$ is not uniformly bounded in any neighborhood of (x_0,u_0). Then there exists sequences $\{(x_\nu,u_\nu)\}$ and $\{\Lambda_\nu\} \subset \mathcal{K}^*$ converging to (x_0,u_0) and ∞ respectively such that

$$\|\Psi(x_\nu,u_\nu,\xi)\bullet\Lambda_\nu\|_\infty \le 1$$

or equivalently

$$-1 \le \Psi(x_\nu,u_\nu,\xi)\bullet\Lambda_\nu \le 1,\ \forall \xi \in \Xi. \tag{2.25}$$

Since $\{\Lambda_\nu/\|\Lambda_\nu\|\}$ is bounded, we may assume without loss of generality that $\Lambda_\nu/\|\Lambda_\nu\| \to \hat{\Lambda}$ as $\nu \to \infty$. Multiplying both sides of equation (2.25) by $1/\|\Lambda_\nu\|$ and driving ν to ∞, we obtain that for each $\xi \in \Xi$,

$$0 \le \lim_{\nu\to\infty} \Psi(x_\nu,u_\nu,\xi)\bullet\frac{\Lambda_\nu}{\|\Lambda_\nu\|} = \Psi(x_0,u_0,\xi)\bullet\hat{\Lambda} \le 0. \tag{2.26}$$

Through the Slater type condition (2.12), the inequality entails that $\hat{\Lambda} = 0$, a contradiction to the fact that $\|\hat{\Lambda}\| = 1$. ■

It is easy to observe that Lemma 2.1 reduces to classical Hoffman's lemma when Ξ is a finite discrete set. Our interest here is in the case when Ξ is a continuous set.

Remark 2.1 Condition (2.12) is a kind of Slater constraint qualification which is widely used in the literature of distributionally robust optimization. It means the underlying functions Ψ in the definition of the ambiguity set through moment conditions cannot be arbitrary. We have a few comments in sequel.

(i) Condition (2.12) is equivalent to

$$0 \in \text{int}\,\{(1-\langle P,1\rangle, \langle P,\Psi(x_0,u_0,\xi)\rangle - \mathcal{K}) : P \in \mathscr{M}_+\}. \tag{2.27}$$

The latter is first considered by Shapiro [19] (see condition (3.12) there) for deriving strong Lagrange duality of moment problems and has been widely used in the literature of distributionally robust optimization, see some detailed analysis of the condition in a recent paper by Xu, Liu and Sun [23]. To see the equivalence, let $\mathcal{U}$ denote an open neighborhood of $(0,0)$ such that

$$\mathcal{U} \subset \text{int}\,\{(1-\langle P,1\rangle, \langle P,\Psi(x_0,u_0,\xi)\rangle - \mathcal{K}) : P \in \mathscr{M}_+\},$$

let $P_0 \in \mathscr{M}_+$ and $\eta_0 \in \mathcal{K}$ be such that $1 - \langle P_0, 1\rangle = 0$ and $0 = \langle P_0, \Psi(x_0, u_0, \xi)\rangle - \eta_0$. Let

$$\mathcal{V} := \{P \in \mathscr{M}_+ : (1 - \langle P, 1\rangle, \langle P, \Psi(x_0, u_0, \xi)\rangle - \mathcal{K}) \in \mathcal{U}\}.$$

Then

$$\begin{array}{rcl} 0 & = & \langle P_0, \Psi(x_0, u_0, \xi)\rangle - \eta_0 \\ & \in & (\langle P, \Psi(x_0, u_0, \xi)\rangle - \mathcal{K}) : P \in \mathcal{V}, \langle P, 1\rangle = 1\} \\ & \subset & \text{int}\, \{(\langle P, \Psi(x_0, u_0, \xi)\rangle - \mathcal{K}) : P \in \mathscr{P}\}. \end{array}$$

Conversely if (2.12) holds, then

$$\begin{array}{rcl} (0,0) & \in & \text{int}\, \{(1 - \langle P, 1\rangle, \langle P, \Psi(x_0, u_0, \xi)\rangle - \mathcal{K}) : P \in \bigcup_{t\in(1-\delta,1+\delta)} t\mathscr{P}\} \\ & \subset & \text{int}\, \{(1 - \langle P, 1\rangle, \langle P, \Psi(x_0, u_0, \xi)\rangle - \mathcal{K}) : P \in \mathscr{M}_+\}. \end{array}$$

(ii) If $\mathcal{K}$ has non-empty interior, then it follows by [2, Proposition 2.106] that condition (2.12) is equivalent to existence of $P_0 \in \mathscr{P}$ such that

$$\langle P_0, \Psi(x_0, u_0, \xi)\rangle \in \text{int}\, \mathcal{K}, \tag{2.28}$$

which is the Slater condition.

(iii) Condition (2.12) implies

$$\{\Lambda \in \mathcal{K}^* : \Psi(x_0, u_0, \xi) \bullet \Lambda \le 0, \forall \xi \in \Xi\} = \{0\}, \tag{2.29}$$

where $\mathcal{K}^*$ is defined in (2.15). Indeed, if this is not true, then there exists a nonzero $\tilde{\Lambda} \in \mathcal{K}^*$ such that

$$\Psi(x_0, u_0, \xi) \bullet \tilde{\Lambda} \le 0, \forall \xi \in \Xi,$$

which implies that

$$\langle P, \Psi(x_0, u_0, \xi)\rangle \bullet \tilde{\Lambda} \le 0, \forall P \in \mathscr{P}. \tag{2.30}$$

On the other hand, under condition (2.12), there exists an open neighborhood $\mathcal{W}$ of 0 such that

$$\mathcal{W} \subset \{(\langle P, \Psi(x_0, u_0, \xi)\rangle - \mathcal{K}) : P \in \mathscr{P}\}.$$

Since $\tilde{\Lambda} \in \mathcal{K}^*\backslash\{0\}$, there exists $\zeta_0 \in \mathcal{W}$ such that $\tilde{\Lambda} \bullet \zeta_0 > 0$. For the given ζ_0, there exist $P_0 \in \mathscr{P}, \eta_0 \in \mathcal{K}$ such that

$$\zeta_0 = \langle P_0, \Psi(x_0, u_0, \xi)\rangle - \eta_0. \tag{2.31}$$

Combining (2.30) and (2.31), we arrive at $(\zeta_0 + \eta_0) \bullet \tilde{\Lambda} \le 0$, which leads to a contraction to $\tilde{\Lambda} \bullet \zeta_0 > 0$ because $\eta_0 \bullet \tilde{\Lambda} \ge 0$.

Condition (2.29) is a kind of Mangsarian-Fromovitz constraint qualification (MFCQ) for the semi-infinite system. In the case when Ψ is continuous, it implies

$$\{\Lambda \in \mathcal{K}^* : \Psi(x, u, \xi) \bullet \Lambda = 0, \forall \xi \in \Xi\} = \{0\}$$

for all $(x, u) \in X \times U$ close to (x_0, u_0).

2.3 Hölder continuity of the ambiguity set

With Lemma 2.1, we are ready to discuss continuity of the set-valued mapping $\mathcal{P}(x,u)$. To simplify the discussion, we confine ourself throughout this subsection and Section 3 to the case that $\Psi(x,u,\xi)$ takes a specific form unless specified otherwise:

$$\Psi(x,u,\xi) = A(\xi) + B(x,u), \tag{2.32}$$

where $A : \mathbb{R}^k \to Y$ and $B : X \times U \to Y$ are single valued mappings. Let $u_0 \in U$ be a fixed parameter value of u and x_0 be any fixed point in X. Our analysis requires B to be Hölder continuous near (x_0,u_0) as formally stated in the following assumption.

Assumption 2.1 *Let $(x_0,u_0) \in X \times U$ be fixed. $B(x,u)$ is Hölder continuous in (x,u) at (x_0,u_0), i.e., there exist $\gamma \in \mathbb{R}_+$ and some positive constants $\nu_1,\nu_2 \in (0,1)$ such that*

$$\|B(x,u) - B(x_0,u_0)\| \le \gamma(\|x-x_0\|^{\nu_1} + \|u-u_0\|^{\nu_2})$$

for $(x,u) \in X \times U$ close to (x_0,u_0).

Theorem 2.1 (Hölder continuity of the ambiguity set mapping) *Assume that the Slater type condition*

$$-B(x_0,u_0) \in \text{int}\,\{(\langle P, A(\xi)\rangle - \mathcal{K}) : P \in \mathscr{P}\} \tag{2.33}$$

holds. Under Assumption 2.1, there exist positive constants C and $\nu_1,\nu_2 \in (0,1)$ such that

$$\mathbb{H}_{TV}(\mathcal{P}(x,u), \mathcal{P}(x',u')) \le C(\|x-x'\|^{\nu_1} + \|u-u'\|^{\nu_2}) \tag{2.34}$$

for any $(x,u),(x',u') \in X \times U$ close to (x_0,u_0).

Proof. Under condition (2.33), we know from Lemma 2.1 that there exists a constant $C_0 > 0$ such that

$$d_{TV}(Q, \mathcal{P}(x,u)) \le C_0 d(\mathbb{E}_Q[\Psi(x,u,\xi)], \mathcal{K}). \tag{2.35}$$

for any $Q \in \mathscr{P}$ and $(x,u) \in X \times U$ close to (x_0,u_0). On the other hand, For any $(x',u') \in X \times U$ close to (x_0,u_0) and $Q \in \mathcal{P}(x',u')$, $d(\mathbb{E}_Q[\Psi(x',u',\xi)],\mathcal{K}) = 0$. Taking into account of Assumption 2.1, we have

$$\begin{aligned} d(\mathbb{E}_Q[\Psi(x,u,\xi)],\mathcal{K}) &\le \|\mathbb{E}_Q[\Psi(x,u,\xi)] - \mathbb{E}_Q[\Psi(x',u',\xi)]\| + d(\mathbb{E}_Q[\Psi(x',u',\xi)],\mathcal{K}) \\ &\le \gamma(\|x-x'\|^{\nu_1} + \|u-u'\|^{\nu_2}). \end{aligned} \tag{2.36}$$

Combining the last inequality with (2.35), we obtain

$$d_{TV}(Q, \mathcal{P}(x,u)) \le C_0\gamma(\|x-x'\|^{\nu_1} + \|u-u'\|^{\nu_2})$$

for any $Q \in \mathcal{P}(x',u')$ and hence

$$\mathbb{D}_{TV}(\mathcal{P}(x,u), \mathcal{P}(x',u')) \le C_0\gamma(\|x-x'\|^{\nu_1} + \|u-u'\|^{\nu_2}). \tag{2.37}$$

Exchanging the positions between (x,u) and (x',u'), we deduce

$$\mathbb{D}_{TV}(\mathcal{P}(x,u), \mathcal{P}(x',u')) \le C_0\gamma(\|x-x'\|^{\nu_1} + \|u-u'\|^{\nu_2}). \tag{2.38}$$

The rest follows from the definition of $\mathbb{H}_{TV}$. ■

It might be helpful to clarify how the specific structure of $\Psi(x, u, \xi)$ defined in (2.32) is exploited in the proof of the theorem. We refer readers to (2.36) where the particular structure gives rise to

$$\|\mathbb{E}_Q[\Psi(x, u, \xi)] - \mathbb{E}_Q[\Psi(x', u', \xi)]\| = \|\mathbb{E}_Q[B(x, u)] - \mathbb{E}_Q[B(x', u')]\| = \|B(x, u) - B(x', u')\|.$$

The situation would become more complicated if the term $A(\xi)$ was not cancelled in that we would require additional conditions which effectively bound $\mathbb{E}_Q[A(\xi)]$. Of course, instead of relying on the specific structure of Ψ, we may alternatively make a blanket assumption such that

$$\|\mathbb{E}_Q[\Psi(x, u, \xi)] - \mathbb{E}_Q[\Psi(x', u', \xi)]\| \leq \gamma(\|x - x'\|^{\nu_1} + \|u - u'\|^{\nu_2})$$

for any $Q \in \mathscr{P}$ and $(x, u), (x', u') \in X \times U$ close to (x_0, u_0) with some positive constants $\gamma > 0$ and $\nu_1, \nu_2 \in (0, 1)$. We will come back to this in Section 5.2.

3 Stability analysis of the optimal values

With the quantitative characterization of the set-valued mapping $\mathcal{P}(x, u)$, we are ready to carry out stability analysis for the inner maximization problem (2.7) or its equivalence (2.11) and (P_u) in terms of the optimal values. Let $v(x, u)$ denote the optimal value of the two problems. Our analysis in this section essentially concerns derivation and quantification of the continuity of $v(x, u)$ in (x, u) and ultimately the continuity of the optimal value function

$$\vartheta(u) := \min_{x \in X} \ v(x, u).$$

The analysis may be viewed as a natural extension of stability analysis in stochastic programming where $\mathcal{P}(x, u)$ is a singleton and is independent of x.

3.1 Pseudo metric and weak compactness

For the purpose of the proposed analysis, we need to introduce another metric which is closely related to the objective function $f(x, \xi)$. Consider the set of random functions:

$$\mathscr{G} := \{g(\cdot) := f(x, \cdot) : x \in X\}.$$

For any two probability measures $P, Q \in \mathscr{P}$, let

$$\mathscr{D}(P, Q) := \sup_{g \in \mathscr{G}} |\mathbb{E}_P[g] - \mathbb{E}_Q[g]|. \tag{3.39}$$

Here we implicitly assume that $\mathscr{D}(P, Q) < \infty$. From (3.39), we can see immediately that $\mathscr{D}(P, Q) = 0$ if and only if

$$\mathbb{E}_P[g] = \mathbb{E}_Q[g], \forall g \in \mathscr{G},$$

which means that convergence of a sequence of probability measures $\{P_N\}$ to P entails uniform convergence of $\mathbb{E}_{P_N}[f(x, \xi)]$ to $\mathbb{E}_P[f(x, \xi)]$. This kind of distance has been widely used for stability analysis in stochastic programming and it is known as *pseudometric* in that it satisfies all properties of a metric except that $\mathscr{D}(Q, P) = 0$ does not necessarily imply $P = Q$ unless the set of functions $\mathscr{G}$ is sufficiently large. For a comprehensive discussion of the concept and related issues, see [18, Sections 2.1-2.2].

Let $Q \in \mathscr{P}$ be a probability measure and $\mathcal{A}_i \subset \mathscr{P}, i = 1, 2$, be a set of probability measures. With the pseudometric, we may define the distance from a single probability measure Q to a set of probability measures $\mathcal{A}_1$ as $\mathscr{D}(Q, \mathcal{A}_1) := \inf_{P \in \mathcal{A}_1} \mathscr{D}(Q, P)$, the deviation (excess) of $\mathcal{A}_1$ from (over) $\mathcal{A}_2$

$$\mathscr{D}(\mathcal{A}_1, \mathcal{A}_2) := \sup_{Q \in \mathcal{A}_1} \mathscr{D}(Q, \mathcal{A}_2)$$

and Hausdorff distance between $\mathcal{A}_1$ and $\mathcal{A}_2$

$$\mathscr{H}(\mathcal{A}_1, \mathcal{A}_2) := \max \left\{ \sup_{P \in \mathcal{A}_1} \mathscr{D}(Q, \mathcal{A}_2), \sup_{P \in \mathcal{A}_2} \mathscr{D}(Q, \mathcal{A}_1) \right\}.$$

In the case when $\mathscr{G}$ is bounded, that is, there exists a positive number M such that $\sup_{g \in \mathscr{G}} \|g\| \leq M$, we are able to establish a relationship between the pseudo metric and the total variation metric that is introduced in the previous section by setting $\tilde{\mathscr{G}} = \mathscr{G}/M$ and consequently

$$\mathscr{D}(P, Q) := M \sup_{\tilde{g} \in \tilde{\mathscr{G}}} |\mathbb{E}_P[\tilde{g}] - \mathbb{E}_Q[\tilde{g}]| \leq M d_{TV}(P, Q).$$

Let $\{P_N\} \in \mathscr{P}$ be a sequence of probability measures. Recall that $\{P_N\}$ is said to converge to $P \in \mathscr{P}$ weakly if

$$\lim_{N \to \infty} \int_\Xi h(\xi) P_N(d\xi) = \int_\Xi h(\xi) P(d\xi), \tag{3.40}$$

for each bounded and continuous function $h : \Xi \to \mathbb{R}$.

Let $\mathcal{A}$ be a set of probability measures on $(\Xi, \mathscr{B})$. Recall that $\mathcal{A}$ is said to be *tight* if for any $\epsilon > 0$, there exists a compact set $\Xi_\epsilon \subset \Xi$ such that $\inf_{P \in \mathcal{A}} P(\Xi_\epsilon) > 1 - \epsilon$. In the case when $\mathcal{A}$ is a singleton, it reduces to the tightness of a single probability measure. $\mathcal{A}$ is said to be *closed* (under the weak topology) if for any sequence $\{P_N\} \subset \mathcal{A}$ with $P_N \to P$ weakly, we have $P \in \mathcal{A}$. $\mathcal{A}$ is said to be *weakly compact* if it is closed under weak topology and bounded.

By the well known Prokhorov's theorem (see [1]), a closed set $\mathcal{A}$ (under the weak topology) of probability measures is *compact* if it is tight. In particular, if Ξ is a compact metric space, then the set of all probability measures on $(\Xi, \mathscr{B})$ is compact in that Ξ is in a finite dimensional space; see [19].

3.2 Hölder continuity of $v(x, u)$ and $\vartheta(u)$

Let $(x_0, u_0) \in X \times U$ be fixed. We make the following assumptions.

Assumption 3.1 *Let $f(x, \xi)$ be defined as in (1.1) and $\hat{\mathcal{P}}$ be a set of probability measures such that $\mathcal{P}(x, u) \subset \hat{\mathcal{P}}$ for (x, u) close to (x_0, u_0). Moreover, the following hold.*

- *(a) For each fixed $\xi \in \Xi$, $f(\cdot, \xi)$ is Lipschitz continuous on X with Lipschitz modulus being bounded by $\kappa(\xi)$, where $\sup_{P \in \hat{\mathcal{P}}} \mathbb{E}_P[\kappa(\xi)] < \infty$.*
- *(b) $\sup_{P \in \hat{\mathcal{P}}} \|\mathbb{E}_P[f(x_0, \xi)]\| < \infty$.*
- *(c) X is a compact set.*

Assumption 3.1 (a) and (b) provide sufficient conditions for the well definedness of the worst expected values of f at any point $x \in X$. Under the assumption, we can easily verify that the pseudometric defined in (3.39) is bounded. The compactness condition on X is imposed for simplicity of analysis, we may weaken the condition to closedness but would then require some inf-compactness conditions of the objective function.

Assumption 3.2 *There exists a weakly compact set $\hat{\mathcal{P}} \subset \mathscr{P}$ such that $\mathcal{P}(x,u) \subset \hat{\mathcal{P}}$ for all (x,u) close to (x_0,u_0)*

The condition is first used in [21], see [21, Assumption 2] and well discussed in [21, Remark 3 and Proposition 7] for moment problems. It provides a sufficient condition for $\sup_{P\in\hat{\mathcal{P}}} \mathbb{E}_P[\kappa(\xi)] < \infty$ and $\sup_{P\in\hat{\mathcal{P}}} \|\mathbb{E}_P[f(x,\xi)]\| < \infty$ for any $x \in X$ by excluding some probability measures with heavy tail. The following proposition can be established analogous to [21, Proposition 2].

Proposition 3.1 *Under Assumptions 3.1, the following assertions hold.*

(a) $\mathbb{E}_P[f(x,\xi)]$ is Lipschitz continuous w.r.t. (P,x) on $\hat{\mathcal{P}} \times X$, that is,

$$|\mathbb{E}_P[f(x,\xi)] - \mathbb{E}_Q[f(y,\xi)]| \leq \mathscr{D}(P,Q) + \sup_{P\in\hat{\mathcal{P}}} \mathbb{E}_P[\kappa(\xi)]\|x-y\|$$

for $P,Q \in \hat{\mathcal{P}}$ and $x,y \in X$.

(b) If, in addition, Assumption 3.2 holds and $\{Pf^{-1}(x,\cdot), P \in \hat{\mathcal{P}}\}$ is uniformly integrable for all x close to x_0, i.e.,

$$\lim_{r\to\infty} \sup_{P\in\hat{\mathcal{P}}} \int_{\{\xi\in\Xi:|f(x,\xi)|\geq r\}} |f(x,\xi)|P(d\xi) = 0,$$

then $v(\cdot,u)$ is equi-Lipschitz continuous on X with modulus being bounded by $\sup_{P\in\hat{\mathcal{P}}} \mathbb{E}_P[\kappa(\xi)]$, that is,

$$|v(x',u) - v(x'',u)| \leq \sup_{P\in\hat{\mathcal{P}}} \mathbb{E}_P[\kappa(\xi)]\|x'-x''\|, \forall x',x'' \in X \tag{3.41}$$

for u close to u_0.

We next investigate stability of the parametric program $(P_{x,u})$ by considering a perturbation of parameter (x,u) in a neighborhood of (x_0,u_0) and quantifying its impact on the optimal value.

Theorem 3.1 (Hölder continuity of $v(x,u)$) *Let $\Psi(x,u,\xi)$ be defined as in (2.32). Let $(x_0,u_0) \in X \times U$ be fixed. Assume the setting and conditions of Theorem 2.1. Assume further: (a) Assumptions 3.1 and 3.2 hold and $\{Pf^{-1}(x,\cdot), P \in \hat{\mathcal{P}}\}$ is uniformly integrable for x close to x_0, (b) $f(x,\xi)$ is bounded, i.e., there exists a positive constant M such that $|f(x,\xi)| \leq M$ for all $(x,\xi) \in X \times \Xi$. Then there exist positive constants $C > 0$ and $\nu_1,\nu_2 \in (0,1)$ such that*

$$|v(x,u) - v(x',u')| \leq C(\|x-x'\|^{\nu_1} + \|u-u'\|^{\nu_2}) \tag{3.42}$$

for $(x,u),(x',u')$ close to (x_0,u_0).

Proof. Let $\mathscr{V}(x,u) := \{\mathbb{E}_P[f(x,\xi)] : P \in \mathcal{P}(x,u)\}$ for fixed $(x,u) \in X \times U$. Under condition (a), $\mathscr{V}(x,u)$ is bounded set in $\mathbb{R}$. Let "conv" denote the convex hull of a set. Then the Hausdorff distance between $\text{conv}\mathscr{V}(x,u)$ and $\text{conv}\mathscr{V}(x_0,u_0)$ can be written as follows:

$$\mathbb{H}(\text{conv}\mathscr{V}(x,u), \text{conv}\mathscr{V}(x',u')) = \max\{|b(x,u) - b(x',u')|, |a(x,u) - a(x',u')|\},$$

where

$$b(x,u) := \sup_{v \in \mathscr{V}(x,u)} v \qquad \text{and} \qquad a(x,u) := \inf_{v \in \mathscr{V}(x,u)} v.$$

By the definition and property of the Hausdorff distance,

$$\begin{aligned}\mathbb{H}(\text{conv}\mathscr{V}(x,u), \text{conv}\mathscr{V}(x',u')) &\leq \mathbb{H}(\mathscr{V}(x,u), \mathscr{V}(x',u')) \\ &= \max(\mathbb{D}(\mathscr{V}(x,u), \mathscr{V}(x',u')), \mathbb{D}(\mathscr{V}(x',u'), \mathscr{V}(x,u))),\end{aligned}$$

where

$$\begin{aligned}\mathbb{D}(\mathscr{V}(x',u'), \mathscr{V}(x,u)) &= \max_{v \in \mathscr{V}(x',u')} d(v, \mathscr{V}(x,u)) \\ &= \max_{v' \in \mathscr{V}(x',u')} \min_{v \in \mathscr{V}(x,u)} \|v - v'\| \\ &= \max_{P \in \mathcal{P}(x',u')} \min_{Q \in \mathcal{P}(x,u)} |\mathbb{E}_P[f(x,\xi)] - \mathbb{E}_Q[f(x,\xi)]| \\ &\leq \max_{P \in \mathcal{P}(x',u')} \min_{Q \in \mathcal{P}(x,u)} \sup_{x \in X} |\mathbb{E}_P[f(x,\xi)] - \mathbb{E}_Q[f(x,\xi)]| = \mathscr{D}(\mathcal{P}(x',u'), \mathcal{P}(x,u)).\end{aligned}$$

In the same manner, we can show $\mathbb{D}(\mathscr{V}(x,u), \mathscr{V}(x',u')) \leq \mathscr{D}(\mathcal{P}(x,u), \mathcal{P}(x',u'))$. Therefore

$$\mathbb{H}(\text{conv}\mathscr{V}(x,u), \text{conv}\mathscr{V}(x',u')) \leq \mathbb{H}(\mathscr{V}(x,u), \mathscr{V}(x',u')) \leq \mathscr{H}(\mathcal{P}(x',u'), \mathcal{P}(x,u)),$$

which gives rise to

$$|v(x,u) - v(x',u')| = \left| \max_{P \in \mathcal{P}(x,u)} \mathbb{E}_P[f(x,\xi)] - \max_{P \in \mathcal{P}(x',u')} \mathbb{E}_P[f(x,\xi)] \right| \leq \mathscr{H}(\mathcal{P}(x',u'), \mathcal{P}(x,u)). \quad (3.43)$$

On the other hand, under condition (b), there exists a constant $M > 0$ such that

$$\mathscr{H}(\mathcal{P}(x',u'), \mathcal{P}(x,u)) \leq M\mathbb{H}_{TV}(\mathcal{P}(x',u'), \mathcal{P}(x,u)). \quad (3.44)$$

Combining (3.43) and (3.44), we immediately obtain (3.42) by virtue of (2.34) in Theorem 2.1. ■

Next we investigate stability of the (1.1) by considering a perturbation of parameter u in a neighborhood of u_0 and quantifying its impact on optimal value.

Theorem 3.2 (Hölder continuity of $\vartheta(u)$) *Let $u_0 \in U$ be fixed. Suppose for each $x \in X$ that: (a) the Slater type condition* (2.33) *holds at x, (b) Assumptions 2.1, 3.1 and 3.2 are satisfied at x, (c) $\{Pf^{-1}(x,\cdot), P \in \hat{\mathcal{P}}\}$ is uniformly integrable and $f(\cdot,\xi)$ is uniformly bounded, that is, there exists a positive constant M such that $|f(x,\xi)| \leq M$ for all $(x,\xi) \in X \times \Xi$. Then there exist positive constants C and ν such that*

$$|\vartheta(u) - \vartheta(u_0)| \leq C\|u - u_0\|^\nu. \quad (3.45)$$

Proof. Under conditions (a)-(c), it follows by Theorem 3.1 that for each $x \in X$, there exists a positive constant $C_x > 0$, depending on x, and $\nu_1, \nu_2 \in (0,1)$ such that

$$|v(x',u') - v(x'',u'')| \leq C_x(\|x' - x''\|^{\nu_1} + \|u' - u''\|^{\nu_2})$$

for $(x', u'), (x'', u'')$ close to (x, u_0). In particular,

$$|v(x', u) - v(x', u_0)| \leq C_x \|u - u_0\|^{\nu_2}$$

for (x', u) close to (x, u_0). Since X is compact, by applying the finite covering theorem, we deduce that there exists a positive constant C such that

$$|v(x, u) - v(x, u_0)| \leq C \|u - u_0\|^{\nu_2}$$

for all $x \in X$ and u close to u_0. The rest follows from Klatte's earlier stability result [11, Theorem 1]. ■

4 Stability analysis of the optimal solutions

In the preceding section, we have carried out quantitative stability analysis for the minimax DRO (P_u) and its inner maximization problem (2.7) in terms of the optimal values. We are short of stating any property of the optimal solutions of these problems primarily because the optimal solution to the inner maximization problem is a probability measure which is relatively difficult to quantify its change against variation of x and u.

To fill out the gap, we take on the challenge in this section by considering Lagrange dual of the inner maximization problem. This is also partially motivated by the fact that with only a few exceptions [15, 13, 23], a majority of numerical methods for solving minimax distributionally robust optimization with moment constraints are developed by converting the inner maximization problem into a deterministic semi-infinite programming problem and then further as a semi-definite programming problem under some more specific conditions on the structure of the underlying functions and the support set Ξ.

Let $\mathcal{K}^*$ denote the dual cone of $\mathcal{K}$ (see the definition in (2.15). For fixed $\Lambda \in \mathcal{K}^*$ and $\lambda \in \mathbb{R}$, let

$$\mathcal{L}(P, \Lambda, \lambda) := \langle P, f(x, \xi) \rangle + \Lambda \bullet \mathbb{E}_P[\Psi(x, u, \xi)] + \lambda(\langle P, 1 \rangle - 1).$$

Through a simple analysis, the Lagrange dual of problem (2.11) can be written as

$$(D_{x,u}) \qquad \begin{array}{cl} \min\limits_{\lambda \in \mathbb{R}, \Lambda \in \mathcal{K}^*} & -\lambda \\ \text{s.t.} & f(x, \xi) + \Lambda \bullet \Psi(x, u, \xi) + \lambda \leq 0, \forall \xi \in \Xi. \end{array} \tag{4.46}$$

This is a cone constrained semi-infinite programming problem when Ξ is an infinite set. By eliminating variable λ, we can reformulate (4.46) as a minimax robust optimization problem

$$\begin{array}{cl} \min\limits_{\Lambda} & \psi(x, u, \Lambda) := \sup\limits_{\xi \in \Xi} [f(x, \xi) + \Lambda \bullet \Psi(x, u, \xi)] \\ \text{s.t.} & \Lambda \in \mathcal{K}^*. \end{array} \tag{4.47}$$

Under the regularity condition (2.27), it follows by [19, Proposition 3.4] and Remark 2.1 that problems (2.11) and (4.47) do not have a dual gap for (x, u) close to (x_0, u_0). Consequently problem (1.1) can be recast as

$$\begin{array}{cl} \min\limits_{x, \Lambda} & \psi(x, u, \Lambda) \\ \text{s.t.} & \Lambda \in \mathcal{K}^*, \\ & x \in X. \end{array} \tag{4.48}$$

It is interesting to note that if Ψ is independent of x and f is convex in x for every fixed ξ, then $\psi(x,u,\Lambda)$ is convex with respect to (x,Λ) and consequently (4.47) is a nonsmooth convex minimization problem.

In the rest of this section, we will carry out stability analysis for problem (4.47) in terms of the optimal solution. The essence of the analysis is to exploit our earlier stability results for cone constrained parametric programming [24].

We start by showing boundedness of the optimal solutions to problem (4.47) under the Slater type conditions (2.12).

Lemma 4.1 (Uniform boundedness of the optimal solutions to problem (4.47)) *Let $(x_0,u_0) \in X \times U$. Suppose that $\Psi(\cdot,\cdot,\xi)$ is continuous at (x_0,u_0) for each $\xi \in \Xi$. If the Slater type condition (2.12) is satisfied, then the set of optimal solutions to the dual problem (4.46) is bounded uniformly for all (x,u) close to (x_0,u_0).*

Proof. It suffices to show that problem (4.47) satisfies the inf-compactness condition, i.e., there exists α_0 such that

$$\mathcal{W}(\alpha_0,x,u) := \{\Lambda \in \mathcal{K}^* : \psi(x,u,\Lambda) \le \alpha_0\}$$

is a compact set for all (x,u) close to (x_0,u_0). Assume for the sake of a contradiction that this is not true. Then for every $\alpha \in \Re$, there exists a sequence $\{\alpha_k\}$ satisfying $\alpha_k \le \alpha$ and a sequence $\{(x_k,u_k)\}$ such that $(x_k,u_k) \to (x_0,u_0)$, $\Lambda_k \in \mathcal{W}(\alpha_k,x_k,u_k)$ for each k, $\|\Lambda_k\| \to \infty$ as $k \to \infty$. Since $\Lambda_k \in \mathcal{W}(\alpha_k,x_k,u_k)$, then

$$f(x_k,\xi) + \Lambda_k \bullet \Psi(x_k,u_k,\xi) \le \alpha_k \le \alpha,\ \forall \xi \in \Xi.$$

Taking a subsequence if necessary, we may assume without loss of generality that $\Lambda_k/\|\Lambda_k\| \to \widetilde{\Lambda}$, which yields $\|\widetilde{\Lambda}\| = 1$. Dividing both sides of the above inequality by $\|\Lambda_k\|$ and driving k to $+\infty$, we obtain

$$\widetilde{\Lambda} \bullet \Psi(x_0,u_0,\xi) \le 0, \forall\ \xi \in \Xi.$$

By Remark 2.1 (iii), the inequalities above imply $\widetilde{\Lambda} = 0$, a contradiction. ∎

In the case when the dual problem (4.46) has a bounded optimal solution uniformly for all (x,u) close to (x_0,u_0), we discuss stability of the set of optimal solutions to problem (4.47). For fixed $(x,u) \in X \times U$, let $v(x,u)$ and $\mathbb{S}(x,u)$ denote respectively the optimal value and the set of optimal solutions of problem (4.47).

Theorem 4.1 (Stability of the optimal solutions of the inner maximization problem (2.7)) *Let $(x_0,u_0) \in X \times U$ be fixed. Assume: (a) the Slater type condition (2.12) is satisfied, (b) $\Psi(x,u,\xi)$ is Hölder continuous w.r.t (x,u) at (x_0,u_0) uniformly for all $\xi \in \Xi$, i.e., there exist positive constants σ and $\nu \in (0,1)$ such that*

$$\|\Psi(x,u,\xi) - \Psi(x_0,u_0,\xi)\| \le \sigma\|(x,u) - (x_0,u_0)\|^\nu$$

for (x,u) close to (x_0,u_0), (c) $\sup_{\xi\in\Xi}\|\Psi(x_0,u_0,\xi)\| < +\infty$, $\sup_{\xi\in\Xi}|f(x_0,\xi)| < +\infty$ and there exists a nonnegative function $\kappa(\xi)$ such that $\sup_{\xi\in\Xi}\kappa(\xi) < +\infty$ and

$$|f(x,\xi) - f(x_0,\xi)| \le |\kappa(\xi)|\|x - x_0\|$$

for x close to x_0, (d) $\psi(x_0,u_0,\Lambda)$ satisfies first order growth condition at the optimal solution set $\mathbb{S}(x_0,u_0)$, i.e., there exists a positive constant $\alpha>0$ such that

$$|\psi(x_0,u_0,\Lambda)-\vartheta(x_0,u_0)|\geq \alpha d\left(\Lambda,\mathbb{S}(x_0,u_0)\right), \quad \forall\, \Lambda\in\mathcal{K}^*.$$

Then the following assertions hold.

(i) There exists a positive constant ϱ such that

$$\mathbb{D}\left(\mathbb{S}(x,u),\mathbb{S}(x_0,u_0)\right)\leq \varrho\|(x,u)-(x_0,u_0)\|^{\nu} \tag{4.49}$$

for all (x,u) close to (x_0,u_0).

(ii) If, in addition, the second order growth condition holds at $\mathcal{Z}(x,u)$ uniformly for all (x,u) close to (x_0,u_0), then

$$\mathbb{H}\left(\mathbb{S}(x,u),\mathbb{S}(x_0,u_0)\right)\leq \varrho\|(x,u)-(x_0,u_0)\|^{\nu}. \tag{4.50}$$

Proof. We use [24, Theorem 2.1] to prove the claims. It is therefore enough to verify the conditions of the theorem. Observe first that under conditions (a)-(c), it follows by Lemma 4.1 that there exists a convex and compact set $\mathcal{W}$ such that problem (4.47) is equivalent to

$$\begin{array}{ll}\min\limits_{\Lambda} & \psi(x,u,\Lambda)\\ \text{s.t.} & \Lambda\in\mathcal{K}^*\cap\mathcal{W}.\end{array}$$

This effectively restricts the feasible set of solutions to a compact set. The rest of the proof are down to verification of the other conditions of [24, Theorem 2.1]. Let $\gamma:=\sup\limits_{\xi\in\Xi}|\kappa(\xi)|+\sup\limits_{\Lambda\in\mathcal{W}}\sigma\|\Lambda\|$. Under conditions (b) and (c),

$$\begin{aligned}|\psi(x,u,\Lambda)-\psi(x_0,u_0,\Lambda)| &= \left|\sup_{\xi\in\Xi}[f(x,\xi)+\Lambda\bullet\Psi(x,u,\xi)]-\sup_{\xi\in\Xi}[f(x_0,\xi)+\Lambda\bullet\Psi(x_0,u_0,\xi)]\right|\\ &\leq \sup_{\xi\in\Xi}|f(x,\xi)-f(x_0,\xi)+\Lambda\bullet\Psi(x,u,\xi)-\Lambda\bullet\Psi(x_0,u_0,\xi)|\\ &\leq \sup_{\xi\in\Xi}\kappa(\xi)\|x-x_0\|+\sigma\|\Lambda\|\|(x,u)-(x_0,u_0)\|^{\nu}\\ &\leq \gamma\|(x,u)-(x_0,u_0)\|^{\nu}\end{aligned}$$

for every (x,u) close to (x_0,u_0). On the other hand, under condition (b), there exists a constant $L:=\sup_{\xi\in\Xi}\|\Psi(x,u,\xi)\|$ such that

$$\begin{aligned}|\psi(x,u,\Lambda)-\psi(x,u,\Lambda')| &= \left|\sup_{\xi\in\Xi}[f(x,\xi)+\Lambda\bullet\Psi(x,u,\xi)]-\sup_{\xi\in\Xi}[f(x,\xi)+\Lambda'\bullet\Psi(x,u,\xi)]\right|\\ &\leq \sup_{\xi\in\Xi}\|\Psi(x,u,\xi)\|\|\Lambda-\Lambda'\|\\ &= L\|\Lambda-\Lambda'\|.\end{aligned}$$

Therefore all conditions of [24, Theorem 2.1] are satisfied. ■

It might be helpful to make some comments on the conditions imposed on Theorem 4.1. Conditions (a) and (b) are used in Section 3. Condition (c) is obvious when Ξ is a compact set but it requires some clarification in general case. From the proof of the theorem, we can see that the condition is fulfilled if $\sup_{\xi\in\Xi}[f(x,\xi)+\Lambda\bullet\Psi(x,u,\xi)]$ and $\sup_{\xi\in\Xi}[f(x,\xi)+\Lambda'\bullet\Psi(x,u,\xi)]$ achieve the maximum in a compact subset of Ξ. It is possible to derive sufficient conditions for the latter using some recent results in variational analysis, we provide some discussions in the appendix as they might be of interest to some readers. To see how the first order growth condition may be possibly satisfied, we consider a simple example.

Example 4.1 Consider the following DRO problem:

$$\begin{aligned} \min_{x\in\mathbb{R}} \ & \max_{P\in\mathscr{P}} \mathbb{E}_P[x+\xi] \\ \text{s.t.} \ & \begin{pmatrix} x+\mathbb{E}_P[\xi] & u+\mathbb{E}_P\left[\sqrt{1-\xi^2}\right] \\ u+\mathbb{E}_P\left[\sqrt{1-\xi^2}\right] & u+\mathbb{E}_P[\xi] \end{pmatrix} \in \mathcal{K}, \end{aligned} \tag{4.51}$$

where ξ is a random variable with support set $\Xi=[-1,1]$, $\mathcal{K}=S_+^{2\times 2}+\left\{\begin{pmatrix} a & 0 \\ 0 & -a\end{pmatrix}: a\in\mathbb{R}\right\}$. By the definition of the dual cone,

$$\left\{\begin{pmatrix} a & 0 \\ 0 & -a\end{pmatrix}: a\in\mathbb{R}\right\}^* = \left\{\begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{12} & \Lambda_{11}\end{pmatrix}: \Lambda_{11},\Lambda_{12}\in\Re\right\}.$$

Moreover, since $\mathcal{K}$ is closed and $S_+^{2\times 2}$ is a self-dual cone, then by [2, Chapter 2],

$$\mathcal{K}^* = (S_+^{2\times 2})^* \bigcap \left\{\begin{pmatrix} a & 0 \\ 0 & -a\end{pmatrix}: a\in\mathbb{R}\right\}^* = \left\{\Lambda\in S^{2\times 2}: \Lambda=\begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{12} & \Lambda_{11}\end{pmatrix}, \Lambda_{11}\geq|\Lambda_{12}|\right\}.$$

Consider $x_0=0, u_0=1$. Through simple calculations, we can derive the Lagrange dual of the inner maximization problem of (4.51)

$$\begin{aligned} \min_{\Lambda} \ & \psi(x_0,u_0,\Lambda) := \sup_{\xi\in[-1,1]} [f(x_0,\xi)+\Lambda\bullet\Psi(x_0,u_0,\xi)] \\ \text{s.t.} \ & \Lambda=\begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{12} & \Lambda_{22}\end{pmatrix}\in\mathcal{K}^*. \end{aligned}$$

In what follows we derive a closed form for the optimal value function $\psi(x_0,u_0,\Lambda)$. Let

$$F(\xi,\Lambda_{11},\Lambda_{12}) := \xi+2\xi\Lambda_{11}+2\sqrt{1-\xi^2}\Lambda_{12}.$$

The maximum of F is achieved either at the boundary of $[-1,1]$ or some stationary point in the interior of the interval. Since $\Lambda_{11}\geq|\Lambda_{12}|$, Λ_{11} can only take non-negative values. The stationary point can be expressed as

$$\hat{\xi} = \begin{cases} \frac{-(2\Lambda_{11}+1)}{\sqrt{(2\Lambda_{11}+1)^2+4\Lambda_{12}^2}}, & \Lambda_{12}\leq 0, \\ \frac{2\Lambda_{11}+1}{\sqrt{(2\Lambda_{11}+1)^2+4\Lambda_{12}^2}}, & \text{otherwise,} \end{cases}$$

with corresponding function values

$$F(\hat{\xi},\Lambda_{11},\Lambda_{12}) = \begin{cases} -\sqrt{(2\Lambda_{11}+1)^2+4\Lambda_{12}^2}, & \Lambda_{12}\leq 0, \\ \sqrt{(2\Lambda_{11}+1)^2+4\Lambda_{12}^2}, & \text{otherwise.} \end{cases}$$

Combining with the function values at the boundary, namely $F(1,\Lambda_{11},\Lambda_{12})=1+2\Lambda_{11}, F(-1,\Lambda_{11},\Lambda_{12})=-1-2\Lambda_{11}$, we obtain

$$\psi(x_0,u_0,\Lambda) = \begin{cases} 1+2\Lambda_{11}, & \Lambda_{12}\leq 0, \\ \sqrt{(2\Lambda_{11}+1)^2+4\Lambda_{12}^2}, & \text{otherwise.} \end{cases}$$

From the expression above, we can see clearly that $\psi(x_0, u_0, \Lambda)$ achieves its minimum when $\Lambda_{11} = 0$. The constraint $\Lambda_{11} \geq |\Lambda_{12}|$ forces $\Lambda_{12} = 0$ giving the optimal solution

$$\mathbb{S}(x_0, u_0) = \left\{ \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \right\}$$

and corresponding optimal value 1. The first order growth condition can be consequently verified in that

$$\begin{aligned}
\psi(x_0, u_0, \Lambda) - 1 &\geq 2\Lambda_{11} + 1 - 1 = 2\Lambda_{11} \\
&\geq \sqrt{2\Lambda_{11}^2 + 2\Lambda_{12}^2} \quad (\text{since } \Lambda_{11} \geq |\Lambda_{12}|) \\
&= \left\| \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{12} & \Lambda_{11} \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \right\|_F \\
&= d\left(\begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{12} & \Lambda_{11} \end{pmatrix}, \mathbb{S}(x_0, u_0) \right)
\end{aligned}$$

for $\Lambda \in \mathcal{K}^*$.

If the Slater condition (2.12) holds at the set of optimal solutions to problem (1.1), then we can convert problem (1.1) into

$$\begin{array}{ll} \min\limits_{x,\Lambda} & \psi(x, u, \Lambda) \\ \text{s.t.} & \Lambda \in \mathcal{K}^*, \\ & x \in X. \end{array} \tag{4.52}$$

In the rest of this section, we will discuss stability of the optimal solution of problem (4.52).

First, let us show that problem (4.52) has a bounded optimal solution set if the Slater type constraint qualification (2.12) holds at any $x_0 \in X$.

Lemma 4.2 *Let $u_0 \in U$ and X be a compact set. Suppose $f(\cdot, \xi)$ and $\Psi(\cdot, \cdot, \xi)$ are continuous on X and $X \times \{u_0\}$ respectively for each $\xi \in \Xi$. If the Slater constraint qualification (2.12) is satisfied for any $x_0 \in X$, then the problem (4.52) has an optimal solution set which is bounded uniformly for all u close to u_0.*

The proof is similar to that of Lemma 4.1, we omit the details here. ■

In the case when problem (4.52) has a bounded optimal solution uniformly for all u close to u_0, we are able to discuss stability of the optimal solutions to problem (4.52). For fixed $u \in U$, let $\vartheta(u)$ and $\mathbb{S}(u)$ denote respectively the optimal value and optimal solution set of problem (4.52) respectively.

Theorem 4.2 (Stability of the optimal solutions of (P_u)) *Let $u_0 \in U$ and X be a compact convex set. Assume: (a) the Slater constraint qualification (2.12) holds at any $x_0 \in X$, (b) there exists $x_0 \in X$ such that $\sup_{\xi \in \Xi} \|\Psi(x_0, u_0, \xi)\| < +\infty$ and there exist $\sigma > 0$ and $\nu \in (0, 1)$ such that*

$$\|\Psi(x, u, \xi) - \Psi(x', u_0, \xi)\| \leq \sigma \|(x, u) - (x', u_0)\|^\nu$$

for u close to u_0 and $x, x' \in X$, (c) there exists $\kappa(\xi)$ satisfying $\sup_{\xi \in \Xi} |\kappa(\xi)| < +\infty$ such that

$$|f(x, \xi) - f(x', \xi)| \leq |\kappa(\xi)| \|x - x'\|$$

for $x, x' \in X$, (d) $\psi(x, u_0, \Lambda)$ satisfies the first order growth condition at the optimal solution set $\mathbb{S}(u_0)$, i.e., there exists a positive constant $\alpha > 0$ such that

$$|\psi(x, u_0, \Lambda) - \vartheta(u_0)| \geq \alpha d\left((x, \Lambda), \mathbb{S}(u_0)\right), \quad \forall\ (x, \Lambda) \in X \times \mathcal{K}^*. \tag{4.53}$$

Then

(i) there exists a positive constant c such that

$$\mathbb{D}\left(\mathbb{S}(u), \mathbb{S}(u_0)\right) \leq c\|u - u_0\|^\nu \tag{4.54}$$

for all u close to u_0;

(ii) if, in addition, the first order growth condition holds at $\mathbb{S}(u)$ uniformly for all u close to u_0, then

$$\mathbb{H}\left(\mathbb{S}(u), \mathbb{S}(u_0)\right) \leq c\|u - u_0\|^\nu. \tag{4.55}$$

Proof. We use [24, Theorem 2.1] to prove the results. It suffices to verify the conditions of the theorem. Under conditions (a)-(c), it follows by Lemma 4.2 that there exists a compact and convex set $\mathcal{W}$ such that problem (4.52) is equivalent to

$$\begin{array}{ll} \min\limits_{x,\Lambda} & \psi(x, u, \Lambda) \\ \text{s.t.} & \Lambda \in \mathcal{K}^* \cap \mathcal{W}, \\ & x \in X. \end{array} \tag{4.56}$$

Consequently, we may analyse stability of the optimal solutions of problem (4.56) whose feasible set of solutions is compact. Under condition (b), we obtain

$$\begin{aligned} |\psi(x, u, \Lambda) - \psi(x, u_0, \Lambda)| &= \left|\sup_{\xi\in\Xi}[f(x, \xi) + \Lambda \bullet \Psi(x, u, \xi)] - \sup_{\xi\in\Xi}[f(x, \xi) + \Lambda \bullet \Psi(x, u_0, \xi)]\right| \\ &\leq \sup_{\xi\in\Xi} |f(x, \xi) - f(x, \xi) + \Lambda \bullet \Psi(x, u, \xi) - \Lambda \bullet \Psi(x, u_0, \xi)| \\ &\leq \sigma\|\Lambda\|\|u - u_0\|^\nu \\ &\leq \gamma\|u - u_0\|^\nu \end{aligned}$$

for u close to u_0, where $\gamma := \sup\limits_{\Lambda\in\mathcal{W}} \sigma\|\Lambda\|$. Moreover, under conditions (b) and (c), there exists a constant $\rho > 0$ such that

$$\begin{aligned} |\psi(x, u, \Lambda) - \psi(x', u, \Lambda')| &= \left|\sup_{\xi\in\Xi}[f(x, \xi) + \Lambda \bullet \Psi(x, u, \xi)] - \sup_{\xi\in\Xi}[f(x', \xi) + \Lambda' \bullet \Psi(x', u, \xi)]\right| \\ &\leq \sup_{\xi\in\Xi} \left|f(x, \xi) - f(x', \xi) + \Lambda \bullet \Psi(x, u, \xi) - \Lambda' \bullet \Psi(x', u, \xi)\right| \\ &\leq \sup_{\xi\in\Xi} |\kappa(\xi)|\|x - x'\| + \sup_{\xi\in\Xi} \|\Psi(x, u, \xi)\|\|\Lambda - \Lambda'\| + L\|\Lambda'\|\|x - x'\|. \end{aligned}$$

Therefore all conditions of [24, Theorem 2.1] are satisfied. ■

The growth condition (4.53) is essential in deriving the stability result. It is difficult to discuss the condition in a general setting. We will revisit the issue under some special circumstance in Section 5.1 where we use directional derivative to characterize the growth condition.

5 Applications

In this section, we apply the stability results established in the preceding sections to a robust formulation of program (1.3) and program (1.5). This will enable us better understand the theoretical results and take a scrutiny of the two important DRO problems.

5.1 Robust one stage stochastic program

We start with program (1.3). Assume that the true probability distribution $P(x, \xi)$ is unknown but it is possible to obtain the true mean value and covariance of ξ or their approximations. Let $\mu(x)$ and $\Sigma(x)$ denote these quantities respectively, let $\mu_0(x)$ and $\Sigma_0(x)$ denote the true mean and and covariance. We consider the ambiguity set

$$\{P \in \mathscr{P} : \mathbb{E}_P[\xi] = \mu(x); \mathbb{E}_P[(\xi - \mu(x))(\xi - \mu(x)^T] \preceq \Sigma(x)\}. \tag{5.57}$$

Obviously the set is uniquely determined by 3-tuples $(x, \mu(\cdot), \Sigma(\cdot))$.

If we use u to denote $(\mu(\cdot), \Sigma(\cdot))$, then the ambiguity set can be written as

$$\mathcal{P}(x, u) := \{P \in \mathscr{P} : \mathbb{E}_P[\xi] = \mu(x); \mathbb{E}_P[(\xi - \mu(x))(\xi - \mu(x)^T] \preceq \Sigma(x)\}. \tag{5.58}$$

Let $u_0 = (\mu_0(\cdot), \Sigma_0(\cdot))$. We can think of parameter u as continuous functions defined over a compact set X and we carry out our stability analysis as u varies near u_0.

To facilitate discussions, we let u_1, u_1^0 denote $\mu(\cdot), \mu_0(\cdot)$ respectively and u_2, u_2^0 denote $\Sigma(\cdot), \Sigma_0(\cdot)$ respectively. Consequently we may confine u to the Cartesian product of spaces $\mathbb{R}^k$ and $S_+^{k\times k}$ equipped with infinity norm, i.e.,

$$\|u\| := \max\left\{\sup_{x\in X} |\mu_l(x)|, l = 1, 2\cdots, k; \quad \sup_{x\in X} |\Sigma_{ij}(x)|, i, j = 1, 2\cdots, k\right\},$$

where $\mu_l(\cdot)$ and $\Sigma_{ij}(x)$ denote the i-th component of $\mu(x)$ and i-th component of the j-th column of matrix $\Sigma(x)$ respectively. To ease the exposition, we write

$$\Psi(x, u, \xi) := A(\xi) + B(x, u),$$

where $A(\xi) := (\xi, \xi\xi^T)$ and $B(x, u) := (u_1(x), (u_1(x)u_1(x)^T - u_2(x))$. The resulting minimax distributionally robust optimization problem can be written as

$$\min_{x\in X} \sup_{P\in\mathcal{P}(x,u)} \mathbb{E}_P[f(x, \xi(\omega))]. \tag{5.59}$$

In what follows, we apply our stability analysis results established in the preceding sections to program (5.59) when u varies near u_0. Our first technical result to be derived is Lipschitz continuity of the ambiguity set $\mathcal{P}(x, u)$ at (x_0, u_0) under the total variation metric. For this purpose, we need the following condition.

Assumption 5.1 *Let $x_0 \in X$ be fixed. There is an open neighborhood $\mathcal{U}_E$ of $\mu_0(x_0)$ such that*

$$\mathcal{U}_E \subset \text{int}\, \{\mathbb{E}_P[\xi] : P \in \mathscr{P}\}$$

and $P_0 \in \mathcal{P}_E := \{P \in \mathscr{P} : \mathbb{E}_P[\xi] \in \mathcal{U}_E\}$ such that

$$\mathbb{E}_{P_0}[(\xi - \mu_0(x_0))(\xi - \mu_0(x_0))^T] \prec \Sigma_0(x_0). \tag{5.60}$$

To see how the assumption may be possibly satisfied, let us consider a simple case when $\Xi = \mathbb{R}^k$. Since $\mathscr{P}$ consists of all Dirac probability measures over $\mathbb{R}^k$ (induced by ξ), we have $\{\mathbb{E}_P[\xi] : P \in \mathscr{P}\} = \mathbb{R}^k$. Consequently the assumption reduces to existence of $P_0 \in \mathscr{P}$ satisfying (5.60) which is guaranteed by [23, Proposition 2.1].

Theorem 5.1 *Let $\mathcal{P}(x,u)$ be defined as in (5.58) and $(x_0,u_0) \in X \times U$ be fixed. Assume: (a) The true mean value $\mu_0(x)$ and covariance matrix $\Sigma_0(x)$ are Lipschitz continuous near x_0, (b) Assumption 5.1 holds. Then there exists a positive constant C such that*

$$\mathbb{H}_{TV}(\mathcal{P}(x,u), \mathcal{P}(x_0,u_0)) \le C(\|x-x_0\|_\infty + \|u-u_0\|_\infty) \tag{5.61}$$

for (x,u) close to (x_0,u_0).

Proof. We use Theorem 2.1 to prove the result. Therefore it suffices to verify the conditions of the theorem. Since $\mu(\cdot)$ and $\mu_0(\cdot)$ are continuous on X and X is compact, then there exist positive constants C_1 and C_2 such that $\|u_1\|_\infty \le C_1$ and $\|u_1^0\|_\infty \le C_2$. Moreover, the Lipschtz continuity of μ_0 and $\Sigma_0(x)$ at x_0 ensures existence of positive constants C_3 and C_4 such that $\|u_1^0(x) - u_1^0(x_0)\|_\infty \le C_3\|x-x_0\|_\infty$ and $\|u_2^0(x) - u_2^0(x_0)\|_\infty \le C_4\|x-x_0\|_\infty$ for x near x_0. Therefore

$$\begin{aligned}
\|B(x,u) - B(x_0,u_0)\| &\le \|u_1(x) - u_1^0(x_0)\|_\infty + \|u_1(x)u_1(x)^T - u_1^0(x_0)u_1^0(x_0)^T\|_\infty + \|u_2(x) - u_2^0(x_0)\|_\infty \\
&\le \|u_1(x) - u_1^0(x)\|_\infty + \|u_1^0(x) - u_1^0(x_0)\|_\infty + \|u_1(x)u_1(x)^T - u_1^0(x)u_1^0(x)^T\|_\infty \\
&\quad + \|u_1^0(x)u_1^0(x)^T - u_1^0(x_0)u_1^0(x_0)^T\|_\infty + \|u_2(x) - u_2^0(x)\|_\infty + \|u_2^0(x) - u_2^0(x_0)\|_\infty \\
&\le \|u-u_0\|_\infty + C_2\|x-x_0\|_\infty + (C_1+C_2)\|u-u_0\|_\infty + 2C_2C_3\|x-x_0\|_\infty \\
&\quad + \|u-u_0\|_\infty + C_4\|x-x_0\|_\infty \\
&\le \max\{2+C_1+C_2, C_2+2C_2C_3+C_4\}(\|x-x_0\|_\infty + \|u-u_0\|_\infty).
\end{aligned}$$

This verifies Assumption 2.1. On the other hand, under Assumption 5.1,

$$(\Sigma_0(x_0) - \mu_0(x_0)\mu_0(x_0)^T) \in \text{int}\ \{\ (\mathbb{E}_P[\xi\xi^T] - S_-^{k\times k}) : P \in \mathcal{P}_E\}$$

and

$$\mu_0(x_0) \in \text{int}\ \{\ \mathbb{E}_P[\xi] : P \in \mathcal{P}_E\}.$$

A combination of these two inclusions gives rise to

$$-B(x_0,u_0) \in \text{int}\ \{\ (\mathbb{E}_P[A(\xi)] - \{0\}_k \times S_-^{k\times k}) : P \in \mathscr{P}\ \},$$

which is the Slater type condition (2.33). ■

From the proof of Theorem 5.1, we can see that Assumption 5.1 entails Slater type condition (2.33). Analogous to the discussions in Section 3.2, the latter enables us to derive the Lagrange dual of the inner maximization problem of program (5.59). To cut short, let $\lambda \in \mathbb{R}^k$ and $\Lambda \in \mathcal{S}_-^k$, let

$$F_u(x,\Lambda,\lambda,\xi) = f(x,\xi) + \lambda^T(\xi - \mu(x)) + \Lambda \bullet (\xi\xi^T - \mu(x)\mu(x)^T - \Sigma(x)), \tag{5.62}$$

where the subscript u denotes $(\mu(\cdot), \Sigma(\cdot))$. Through the Lagrange dual, the inner maximization problem can be written as

$$\begin{aligned}
\min_{\Lambda,\lambda}\quad & \psi_u(x,\Lambda,\lambda) := \sup_{\xi\in\Xi} F_u(x,\Lambda,\lambda,\xi). \\
\text{s.t.}\quad & \Lambda \in S_-^{k\times k}, \\
& \lambda \in \mathbb{R}^m,
\end{aligned} \tag{5.63}$$

and consequently the minimax distributionally robust optimization problem (5.59) can be equivalently written as

$$\begin{array}{ll} \min\limits_{x,\Lambda,\lambda} & \psi_u(x,\Lambda,\lambda) \\ \text{s.t.} & x \in X, \\ & \Lambda \in S_-^{k\times k}, \\ & \lambda \in \mathbb{R}^m. \end{array} \tag{5.64}$$

Next, we proceed to quantitative analysis of problem (5.64) as u varies near u_0. Let $\vartheta(u)$ and $\mathbb{S}(u)$ denote the optimal value and the set of optimal solutions of program (5.64).

Theorem 5.2 *Suppose: (a) Assumption 5.1 holds at every point $x \in X$, (b) the true mean value $\mu_0(x)$ and covariance matrix $\Sigma_0(x)$ are Lipschitz continuous on X and $\sup_{\xi\in\Xi}\|A(\xi)\| < +\infty$ holds, (c) there exists a nonnegative function $\kappa(\xi)$ such that $\sup_{\xi\in\Xi}\kappa(\xi) < +\infty$ and*

$$|f(x,\xi) - f(x',\xi)| \leq \kappa(\xi)\|x - x'\|, \ \forall x, x' \in X,$$

(d) $\psi_{u_0}(\cdot)$ satisfies the first order growth condition at $\mathbb{S}(u_0)$, i.e.,

$$\psi_{u_0}(x,\Lambda,\lambda) - \vartheta(u_0) \geq \alpha d((x,\Lambda,\lambda), \mathbb{S}(u_0)) \tag{5.65}$$

for all $(x,\Lambda,\lambda) \in X \times S_-^{k\times k} \times \mathbb{R}^k$. Then there exists a positive constant C such that

$$\mathbb{D}\left(\mathbb{S}(u), \mathbb{S}(u_0)\right) \leq C\|u - u_0\| \tag{5.66}$$

for u close to u_0.

Proof. We use Theorem 4.2 to prove the claim. Therefore it suffices to verify the conditions of the theorem. Under Assumption 5.1, the Slater type condition (2.12) holds at x_0, therefore condition (a) of Theorem 4.2 is satisfied. Condition (b) ensures the condition (b) of Theorem 4.2. Conditions (c) and (d) coincide with the conditions (c) and (d) of Theorem 4.2. ■

Remark 5.1 In some cases, conditions (b), (c) and (d) of Theorem 5.2 may be weakened. From the proof of Theorem 4.2, we can see that if there exists a set $\Theta_u \subset X \times S_-^{k\times k} \times \mathbb{R}^m$ such that problem (5.64) is equivalent to the following:

$$\begin{array}{ll} \min\limits_{x,\Lambda,\lambda} & \tilde{\psi}_u(x,\Lambda,\lambda) \\ \text{s.t.} & (x,\Lambda,\lambda) \in \Theta_u, \end{array}$$

then conditions (b), (c) and (d) of Theorem 5.2 can be replaced by the following:

(b') $\tilde{\psi}_u(x,\Lambda,\lambda)$ is Hölder continuous w.r.t. u at u_0 for $(x,\Lambda,\lambda) \in \Theta_u$.

(c') $\tilde{\psi}_u(x,\Lambda,\lambda)$ is Lipschitz continuous w.r.t (x,Λ,λ) over Θ_u.

(d') There exist positive constants α and θ such that

$$\tilde{\psi}_{u_0}(x,\Lambda,\lambda) - \vartheta(u_0) \geq \alpha d((x,\Lambda,\lambda), \mathbb{S}(u_0))^\theta$$

for all $(x,\Lambda,\lambda) \in \Theta_{u_0}$ with $\theta = 1$ or 2.

The weakened conditions are more easily satisfied. We explain this through an example.

Example 5.1 Consider DRO problem (5.59) with $f(x,\xi) = x^T\xi$, where $x \in X$ and X is a compact set containing 0, ξ is a random vector with support set $\Xi = \mathbb{R}^k$.

The ambiguity set is defined by (5.58) with $\mu(x) = Bx$ and $\mu_0(x) = B_0x$, where both $B, B_0 \in \text{int}\, \mathcal{S}_+^k$, in other words, we consider variation of $\mu(\cdot)$ near $\mu_0(\cdot)$ via perturbation of matrix B near B_0 in the interior of $\mathcal{S}_+^k$. The upper bound of the covariance matrix $\Sigma(x)$ in the moment condition is a positive definite matrix, i.e., $\Sigma(x) \in \text{int}\, \mathcal{S}_+^k$ for all $x \in X$.

Let $u = (\mu(\cdot), \Sigma(\cdot))$ and $u_0 = (\mu_0(\cdot), \Sigma(\cdot))$ leaving $\Sigma(\cdot)$ unchanged to simplify the discussion.

With these specific details, we can write down the Lagrange dual of the inner maximization problem (5.63) as

$$\begin{array}{ll} \min\limits_{x,\Lambda,\lambda} & \psi_u(x,\Lambda,\lambda) := \sup\limits_{\xi\in\mathbb{R}^k} \left[x^T\xi + \lambda^T(\xi - Bx) + \Lambda \bullet (\xi\xi^T - Bxx^TB - \Sigma(x))\right] \\ \text{s.t.} & (x,\Lambda,\lambda) \in X \times S_-^{k\times k} \times \mathbb{R}^m. \end{array} \tag{5.67}$$

In what follows, we verify the conditions of Theorem 5.2 or their counterparts in Remark 5.1. First, since $\Xi = \mathbb{R}^k$, it follows by the comments after Assumption 5.1, condition (a) holds. Second, we can show that problem (5.67) can be equivalently written as

$$\begin{array}{ll} \min\limits_{x,\Lambda,\lambda} & \tilde{\psi}_u(x,\Lambda,\lambda) := x^TBx \\ \text{s.t.} & x \in X, x = -\lambda, \Lambda = 0. \end{array} \tag{5.68}$$

To see this, we present problem (5.67) in epigraphical form by introducing a new variable t

$$\begin{array}{ll} \min\limits_{t,x,\Lambda,\lambda} & t \\ \text{s.t.} & x^T\xi + \lambda^T(\xi - Bx) + \Lambda \bullet (\xi\xi^T - Bxx^TB - \Sigma(x)) \le t, \quad \forall \xi \in \mathbb{R}^k, \\ & (x,\Lambda,\lambda) \in X \times S_-^{k\times k} \times \mathbb{R}^m. \end{array}$$

Letting $s := -\lambda B^T x - \Lambda \bullet Bxx^TB - \Lambda \bullet \Sigma(x) - t$ and reformulating the semi-infinite constraints as a semi-definite constraint, the program above can be further formulated as

$$\begin{array}{ll} \min\limits_{s,x,\Lambda,\lambda} & \begin{pmatrix} s & \frac{x+\lambda}{2} \\ \frac{(x+\lambda)^T}{2} & \Lambda \end{pmatrix} \bullet \begin{pmatrix} -1 & -Bx \\ -(Bx)^T & -(\Sigma(x) + Bxx^TB) \end{pmatrix} + x^TBx \\ \text{s.t.} & \begin{pmatrix} s & \frac{x+\lambda}{2} \\ \frac{(x+\lambda)^T}{2} & \Lambda \end{pmatrix} \preceq 0, \\ & (x,\Lambda,\lambda) \in X \times S_-^{k\times k} \times \mathbb{R}^m. \end{array} \tag{5.69}$$

Since $\Sigma(x) \succ 0$, and $Bxx^TB \succeq 0$, then

$$\begin{pmatrix} -1 & -Bx \\ -(Bx)^T & -(\Sigma(x) + Bxx^TB) \end{pmatrix} \prec 0.$$

Thus, the linear semi-definite programming problem achieves its minimum when

$$\begin{pmatrix} s & \frac{x+\lambda}{2} \\ \frac{(x+\lambda)^T}{2} & \Lambda \end{pmatrix} = 0,$$

i.e., $x = -\lambda$ and $t = x^T Bx$. This shows formulation (5.68). Let

$$\Theta_u := \{(x, \Lambda, \lambda) : x \in X, \lambda = -x, \Lambda = 0\}.$$

Within Θ_u, $\tilde{\psi}_u(x, \Lambda, \lambda)$ depends on u only through B. Since x is confined to compact set X. Therefore $\tilde{\psi}_u(x, \Lambda, \lambda)$ is globally Lipschitz continuous w.r.t. B and hence Lipschitz continuous in u near u_0 uniformly for all $(x, \Lambda, \lambda) \in \Theta_u$. Moreover, for fixed u (i.e. B), $\tilde{\psi}_u(x, \Lambda, \lambda)$ is Lipschitz continuous w.r.t (x, Λ, λ) on Θ_u because it is a convex quadratic function of x. This verifies conditions (b') and (c') of Remark 5.1. Third, since $0 \in X$, through a direction calculation, we obtain $\vartheta(u_0) = 0$ and

$$\mathbb{S}(u_0) := \left\{(x^*, 0, x^*) : (x^*)^T B_0 x^* = 0, x^* \in X\right\} = \{(0, 0, 0)\}.$$

Therefore there exists constant $\alpha > 0$ such that

$$\tilde{\psi}_{u_0}(x, \Lambda, \lambda) - \vartheta(u_0) = x^T B_0 x - 0 \geq \alpha \|x\|^2 = \frac{\alpha}{2} d((x, \Lambda, \lambda), \mathbb{S}(u_0))^2 \quad \forall (x, \Lambda, \lambda) \in \Theta_{u_0}.$$

This verifies condition (d'). Therefore by Theorem 5.2, there exists a positive constant C such that

$$\mathbb{D}\left(\mathbb{S}(u), \mathbb{S}(u_0)\right) \leq C\|u - u_0\|_\infty \tag{5.70}$$

for u close to u_0.

On the other hand, it is easy to derive that $\vartheta(u) = 0$ and

$$\mathbb{S}(u) := \left\{(x^*, 0, x^*) : (x^*)^T B x^* = 0, x^* \in X\right\} = \{(0, 0, 0)\}$$

for all B close to B_0. Consequently (5.70) may be verified directly

$$0 = \mathbb{D}\left(X \cap \{x \in \mathbb{R}^n : x^T Bx = 0\}, X \cap \{x \in \mathbb{R}^n : x^T B_0 x = 0\}\right) < \sup_{x \in X} \|Bx - B_0 x\| \leq C\|u - u_0\|_\infty$$

for some positive constant C.

5.1.1 The growth condition

The growth condition plays an important role in Theorem 5.2. Here we take a close look at the condition. Specifically, we will use directional derivative to characterize the condition.

Let us consider a generic optimization problem

$$\min_{y \in \mathcal{Y}} \; g(y), \tag{5.71}$$

where $g : Y \to \mathbb{R}$ is a function, Y is a Banach space and $\mathcal{Y} \subset Y$ is a convex and compact set. Let ϑ and S denote the optimal value and the set of optimal solutions respectively. The following result says that the directional derivative function of g implies the first order growth condition.

Lemma 5.1 *Assume: (a) g is Fréchet directionally differentiable on S, (b) there exists a positive constant α such that for any $y^* \in S$,*

$$g'(y^*, d) \geq \alpha\|d\|, \; \forall d \in \mathcal{Y} - y^*. \tag{5.72}$$

Then

(i) the first order growth condition holds at set S locally with rate $\frac{\alpha}{2}$, that is, there exists $\delta > 0$ such that

$$g(y) - \vartheta \geq \frac{\alpha}{2} d(y, S), \forall y \in (S + \delta\mathcal{B}) \cap \mathcal{Y}; \tag{5.73}$$

(ii) if, in addition, the function g is convex, then the first order growth condition (5.73) holds for all $y \in \mathcal{Y}$.

Proof. Part (i). Let $y^* \in S$. Under condition (a), it follows by Fréchet differentiability of φ that

$$g(x) = g(y^*) + g'(y^*, d) + o(\|y - y^*\|)$$

for $y \in \mathcal{Y}$ close to y^*, where we write $o(t)$ for the term such that $o(t)/t \to 0$ as $t \to 0$. Through condition (b), this means that there exists a positive constant δ_{y^*} depending on y^* such that

$$g(y) - g(y^*) \geq \frac{\alpha}{2} d(y, y^*) \geq \frac{\alpha}{2} d(y, S)$$

for any $y \in \mathcal{Y} \cap \operatorname{int} B(y^*, \delta_{y^*})$. Thus S may be covered by the union of a collection of δ-balls, i.e.,

$$S \subseteq \bigcup_{y \in S} \operatorname{int}\, B(y, \delta_y).$$

Since S is a compact set, by the finite covering theorem, there exist a finite number of points $y_1, y_2, \cdots, y_k \in S$ and positive constants $\delta_{y_i}, i = 1, \cdots, k$ and $\delta > 0$ such that

$$S + \delta\mathcal{B} \subseteq \bigcup_{i=1}^{k} \operatorname{int}\, B(y_i, \delta_{y_i}) \tag{5.74}$$

and

$$g(x) - \vartheta \geq \frac{\alpha}{2} d(y, S) \tag{5.75}$$

for any $y \in (S + \delta\mathcal{B}) \cap \mathcal{Y}$. This shows Part (i).

Part (ii). It suffices to show (5.75) holds for any $y \in \mathcal{Y} \backslash (S + \delta\mathcal{B})$. Let

$$y \in \mathcal{Y} \backslash (S + \delta\mathcal{B})$$

and $\hat{y} = \Pi_S(y)$ the orthogonal projection of y on S. Let $\tilde{y} = \Pi_{S+\delta\mathcal{B}}(y')$. By the convexity set S, $\tilde{y}$ lies on the line segment connecting $\hat{y}$ and y, that is, there exists $t \in (0, 1)$ such that $\tilde{y} = ty + (1 - t)\hat{y}$. Since $\mathcal{Y}$ is a convex, all of the three points are contained in $\mathcal{Y}$. By the convexity of $g(\cdot)$ and the local growth condition (5.75),

$$\begin{aligned} tg(y) + (1 - t)\vartheta &= tg(y) + (1 - t)g(\hat{y}) \geq g(\tilde{y}) \\ &\geq \vartheta + \frac{\alpha}{2} d(\tilde{y}, S) \text{ (condition (5.75))} \\ &= \vartheta + \frac{\alpha}{2} \|ty + (1 - t)\hat{y} - \hat{y}\| \\ &= \vartheta + \frac{\alpha}{2} t\|y - \hat{y}\| = \vartheta + \frac{\alpha}{2} td(y, S), \end{aligned}$$

which, through a simple rearrangement, yields

$$g(y) - \vartheta \geq \frac{\alpha}{2} d(y, S). \tag{5.76}$$

Combining (5.75) and (5.76), we obtain the conclusion. ■

We now return to discuss the growth condition of Theorem 5.2 using Lemma 5.1. For fixed $(x_0, \Lambda_0, \lambda_0) \in \mathbb{S}(u_0)$, let $\Xi^*(x_0, \Lambda_0, \lambda_0, u_0)$ denote the set of optimal solutions to the maximization problem

$$\sup_{\xi \in \Xi} F_{u_0}(x_0, \Lambda_0, \lambda_0, \xi) \tag{5.77}$$

and

$$\mathcal{L}_{u_0}(x_0, \Lambda_0, \lambda_0, \xi; d_x, d_\Lambda, d_\lambda) := \nabla_x f(x_0, \xi)^T d_x + \xi^T d_\Lambda \xi + \xi^T d_\lambda + \mathcal{H}_{u_0}(x_0, \Lambda_0, \lambda_0; d_x, d_\Lambda, d_\lambda),$$

where $(d_x, d_\Lambda, d_\lambda) \in \mathbb{R}^n \times S^{k\times k} \times \mathbb{R}^k$ is fixed and

$$\begin{aligned} \mathcal{H}_{u_0}(x_0, \Lambda_0, \lambda_0; d_x, d_\Lambda, d_\lambda) &= \left[2\nabla \mu_0(x_0)^T \Lambda_0 \mu_0(x_0) - \nabla \mu_0(x_0)^T \lambda_0 - \nabla_x(\Sigma_0(x_0) \bullet \Lambda_0)\right]^T d_x \\ &\quad -\mu_0(x_0)^T d_\lambda + \left[\mu_0(x_0)\mu_0(x_0)^T - \Sigma_0(x_0)\right] \bullet d_\Lambda. \end{aligned} \tag{5.78}$$

Assume that problem (5.77) satisfies sup-compactness condition and $F_{u_0}(\cdot, \cdot, \cdot, \xi)$ is continuous differentiable on $X \times S^k_- \times \mathbb{R}^k$ for each $\xi \in \Xi$. Then by the well known Danskin's Theorem (see [6]), ψ_{u_0} is Fréchet directionally differentiable and its derivative can be written as

$$\psi'_{u_0}(x_0, \Lambda_0, \lambda_0; d_x, d_\Lambda, d_\lambda) = \sup_{\xi \in \Xi^*(x_0, \Lambda_0, \lambda_0, u_0)} \mathcal{L}_{u_0}(x_0, \Lambda_0, \lambda_0; \xi, d_x, d_\Lambda, d_\lambda). \tag{5.79}$$

Proposition 5.1 *Let fixed $u_0 \in U$ be fixed. Suppose: (a) $f(\cdot, \xi)$, $\mu_0(\cdot)$ and $\Sigma_0(\cdot)$ are continuously differentiable on X, (b) the sup-compactness condition holds at any $(x_0, \Lambda_0, \lambda_0) \in \mathbb{S}(u_0)$, i.e., for each $(x_0, \Lambda_0, \lambda_0) \in \mathbb{S}(u_0)$, there exists $\alpha \in \mathbb{R}$ and a compact set $\tilde{\Xi} \subset \Xi$ such that*

$$\emptyset \neq \{\xi \in \Xi : F(x, \Lambda, \lambda, u_0, \xi) \geq \alpha\} \subset \tilde{\Xi}$$

for every (x, Λ, λ) close to $(x_0, \Lambda_0, \lambda_0)$, where $F(x, \Lambda, \lambda, u, \xi)$ is defined as (5.62). Then the following assertions hold.

(i) $\psi_{u_0}(x, \Lambda, \lambda)$ is Fréchet directionally differentiable at any $(x_0, \Lambda_0, \lambda_0) \in \mathbb{S}(u_0)$ and its directional derivative is defined as in (5.79).

(ii) If, in addition, $\Xi^*(x_0, \Lambda_0, \lambda_0, u_0) = \{\xi_0\}$, then $\psi(x, \Lambda, \lambda, u_0)$ is Fréchet differentiable at $(x_0, \Lambda_0, \lambda_0)$.

Proof. Note that condition (a) implies the continuous differentiability of $F_{u_0}(x, \Lambda, \lambda, \xi)$ on $X \times S^k_- \times \mathbb{R}^k$ for each $\xi \in \Xi$. Then rest follows by virtue of [2, Theorem 4.13]. ■

From the proof of Theorem 5.1, we know that Assumption 5.1 entails the Slater type condition (2.33), which by Lemma 4.2, means that $\mathbb{S}(u)$ is uniformly bounded for all u close to u_0. Therefore, if Assumption 5.1 holds at every point $x \in X$, then there exists a convex and compact set $\mathcal{W}$ such that problem (5.64) is equivalent to

$$\begin{aligned} \min_{x,\Lambda,\lambda} \quad & \psi_u(x, \Lambda, \lambda) \\ \text{s.t.} \quad & (x, \Lambda, \lambda) \in (X \times S^{k\times k}_- \times \mathbb{R}^m) \cap \mathcal{W} \end{aligned} \tag{5.80}$$

for $u \in U$ close to u_0. Problem (5.80) effectively confines Λ and λ to a compact set.

With Lemma 5.1 and Proposition 5.1, we are ready to characterize the first order growth condition of problem (5.80) in terms of directional derivatives.

Proposition 5.2 *Assume the setting and conditions of Proposition 5.1. If* $\psi_{u_0}(\cdot)$ *is convex on* $X \times S_-^{k\times k} \times \mathbb{R}^k$ *and there exists a positive constant* α *such that for each* $(x_0, \Lambda_0, \lambda_0) \in \mathbb{S}(u_0)$*,*

$$\sup_{\xi\in\Xi^*(x_0,\Lambda_0,\lambda_0,u_0)} \left(\nabla_x f(x_0,\xi)^T d_x + \xi^T d_\Lambda \xi + \xi^T d_\lambda + \mathcal{H}_{u_0}(x_0,\Lambda_0,\lambda_0; d_x, d_\Lambda, d_\lambda)\right) \geq \alpha\|(d_x, d_\Lambda, d_\lambda)\|,$$

for all $(d_x, d_\Lambda, d_\lambda) \in (X \times S_-^{k\times k} \times \mathbb{R}^k) \cap \Theta - (x_0, \Lambda_0, \lambda_0)$*, where* $\mathcal{H}_{u_0}$ *is defined in (5.78), then*

$$\psi_{u_0}(x, \Lambda, \lambda) - \vartheta(u_0) \geq \frac{\alpha}{2} d((x, \Lambda, \lambda), \mathbb{S}(u_0))$$

for all $(x, \Lambda, \lambda) \in (X \times S_-^{k\times k} \times \mathbb{R}^k) \cap \mathcal{W}$*.*

5.2 Robust program (1.5)

We now turn to discuss problem (1.5) which is a robust formulation of one stage stochastic program with expected inequality constraints. Assuming the ambiguity set $\mathcal{P}(u)$ is defined through moment conditions and $g(x, \xi)$ takes a specific form, i.e.,

$$g(x, \xi) = C(\xi)x + d(\xi)$$

where $C(\xi)$ is a random matrix and $d(\xi)$ is a random vector. Then we might combine the moment conditions and the stochastic inequality constraints $g(x, \xi)$ by considering the following robust optimization problem

$$\begin{array}{cl} \min\limits_{x\in X} \sup\limits_{P\in\mathscr{P}} & \mathbb{E}_P[f(x, \xi(\omega))] \\ \text{s.t.} & \mathbb{E}_P[A(\xi)]x \leq b(x, u), \end{array} \tag{5.81}$$

where the inequality constraints capture both the moment conditions and inequality constraints $\mathbb{E}_P[g(x, \xi)] \leq 0$. By slightly abusing the notation, we let

$$g(x, u, \xi) := A(\xi)x - b(x, u),$$

where $A(\cdot) : \mathbb{R}^k \to \mathbb{R}^{m\times n}, b(\cdot, \cdot) : X \times U \to \mathbb{R}^m, f(\cdot, \cdot) : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}$.

Theorem 5.3 *Let* $(x_0, u_0) \in X \times U$ *be fixed. Suppose: (a)* $A(\cdot)$ *is continuous on* Ξ *and* Ξ *is a compact set, moreover there exists* $P_0 \in \mathscr{P}$ *such that*

$$\mathbb{E}_{P_0}[A(\xi)]x < b(x_0, u_0), \tag{5.82}$$

(b) $b(x, u)$ *is Hölder continuous in* (x, u) *at* (x_0, u_0)*, i.e., there exist* $\gamma \in \mathbb{R}_+$ *and some positive constants* $\nu_1, \nu_2 \in (0, 1)$ *such that*

$$\|b(x, u) - b(x_0, u_0)\| \leq \gamma(\|x - x_0\|^{\nu_1} + \|u - u_0\|^{\nu_2})$$

for $(x, u) \in X \times U$ *close to* (x_0, u_0)*. Then there exist positive constants* C*,* ν_1 *and* ν_2 *such that*

$$\mathbb{H}_{TV}(\mathcal{P}(x, u), \mathcal{P}(x_0, u_0)) \leq C(\|x - x_0\|^{\nu_1} + \|u - u_0\|^{\nu_2})$$

for (x, u) *close to* (x_0, u_0)*, where* $\mathcal{P}(x, u) = \{P \in \mathscr{P} : \mathbb{E}_P[g(x, u, \xi)] \leq 0\}$*.*

Proof. Condition (5.82) means that the Slater type condition (2.12) holds. Moreover, since Ξ is a compact set and $A(\cdot)$ is continuous, then

$$\sup_{P\in\mathscr{P}} \mathbb{E}_P[\|A(\xi)\|_F] < +\infty.$$

Furthermore, under condition (b), we have

$$\begin{aligned}\|\mathbb{E}_Q[g(x,u,\xi)] - \mathbb{E}_Q[g(x_0,u_0,\xi)]\| &\leq \|\mathbb{E}_Q[A(\xi)]x - \mathbb{E}_Q[A(\xi)]x_0\| + \|b(x,u) - b(x_0,u_0)\| \\ &\leq \sup_{Q\in\mathscr{P}} \mathbb{E}_Q[\|A(\xi)\|_F]\|x-x_0\| + \gamma(\|x-x_0\|^{\nu_1} + \|u-u_0\|^{\nu_2}) \\ &\leq L(\|x-x_0\|^{\nu_1} + \|u-u_0\|^{\nu_2})\end{aligned}$$

for any $Q \in \mathscr{P}$ and (x,u) close to (x_0,u_0) , where $\|\cdot\|_F$ denotes the Frobenius norm of matrix and

$$L = \sup_{Q\in\mathscr{P}} \mathbb{E}_Q[\|A(\xi)\|_F] + \gamma.$$

The rest follow from Theorem 2.1 and the follow-up comments. ■

Assumption 5.2 *For any $x_0 \in X$, there exists $P_0 \in \mathscr{P}$ such that $\mathbb{E}_{P_0}[A(\xi)]x_0 < b(x_0,u_0)$.*

Under Assumption 5.2, it follows from previous discussions that we can recast the inner maximization problem of program (5.81) through Lagrange dual as

$$\begin{array}{ll}\min\limits_{\lambda} & \psi_u(x,\lambda) \\ \text{s.t.} & \lambda \in \mathbb{R}^m_-,\end{array} \tag{5.83}$$

where $\psi_u(x,\lambda) = \sup\limits_{\xi\in\Xi} F_u(x,\lambda,\xi)$ and

$$F_u(x,\lambda,\xi) = f(x,\xi) + \lambda^T(A(\xi)x - b(x,u)). \tag{5.84}$$

Consequently we can reformulate problem (5.59) as

$$\begin{array}{ll}\min\limits_{x,\lambda} & \psi_u(x,\lambda) \\ \text{s.t.} & x\in X, \lambda \in \mathbb{R}^m_-.\end{array} \tag{5.85}$$

Let $\tilde{\mathbb{S}}(u)$ denote the set of the optimal solutions to problem (5.85), $\Xi^*(x_0,\lambda_0,u_0)$ the optimal solution set of $\sup\limits_{\xi\in\Xi} F_{u_0}(x_0,\lambda_0,\xi)$, and

$$\tilde{\mathcal{L}}_{u_0}(x_0,\lambda_0,\xi,d_x,d_\lambda) = [\nabla_x f(x_0,\xi) + A(\xi)^T\lambda_0 - \nabla_x b(x_0,u_0)]^T d_x + [A(\xi)x_0 - b(x_0,u_0)]^T d_\lambda.$$

The following theorem characterizes stability of the problem.

Theorem 5.4 *Let $u_0 \in U$ and X be a compact and convex set. Suppose (a) Assumption 5.2 holds, (b) $f(\cdot,\xi)$ and $b(\cdot,u_0)$ are continuously differentiable on X, (c) for each $(x_0,\lambda_0) \in \tilde{\mathbb{S}}(u_0)$, $\Xi^*(x,\Lambda,\lambda) \neq \emptyset$ for all (x,λ) close to (x_0,λ_0), (d) $\psi_{u_0}(\cdot)$ is convex on $X \times \mathbb{R}^m_-$ (e) there exists positive measurable function $\kappa(\xi)$ such that $\sup_{\xi\in\Xi}\kappa(\xi) < +\infty$ and*

$$|f(x,\xi) - f(x',\xi)| \leq \kappa(\xi)\|x-x'\|, \forall x,x' \in X,$$

(g) there exists a positive constant α such that for any $(x_0, \lambda_0) \in \tilde{\mathbb{S}}(u_0)$,

$$\sup_{\xi \in \Xi^*(x_0,\lambda_0,u_0)} \tilde{\mathcal{L}}_{u_0}(x_0, \lambda_0, \xi, d_x, d_\lambda) \geq \alpha \|(d_x, d_\lambda)\|,$$

for all $(d_x, d_\lambda) \in (X \times \mathbb{R}^m_-) - (x_0, \lambda_0)$. Then there exists a positive constant c such that

$$\mathbb{D}\left(\tilde{\mathbb{S}}(u), \tilde{\mathbb{S}}(u_0)\right) \leq c\|u - u_0\| \tag{5.86}$$

for all u close to u_0.

Proof. The proof follows directly from Theorem 4.2 and Proposition 5.2. ■

Note that it is possible to investigate stability of problem (1.5) via (1.6) given their equivalence. We leave this to interested readers.

References

[1] K. B. Athreya and S. N. Lahiri, *Measure Theory and Probability Theory*, Springer, NewYork, 2006.

[2] J. F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer-Verlag, New York, 2000.

[3] M. Breton, S. E. Hachem, Algorithms for the solution of stochastic dynamic minimax problems, *Computational Optimization and Applications*, Vol. 4, pp. 317-345, 1995.

[4] M. Breton, S.E. Hachem, A scenario aggregation algorithm for the solution of stochastic dynamic minimax problems, *Stochastics and Stochastic Reports*, Vol. 53, pp. 305-322, 1995.

[5] J. Dupačová, Uncertanities in minimax stochastic programs, *Optimization*, Vol. 60, pp. 10-11, 2011.

[6] J. Danskin, The Theory of Max-rain and Its Applications to Weapons Allocation Problems, Springer, 1967.

[7] Z. Hu and L. J. Hong, Kullback-Leibler divergence constrained distributionally robust optimization, *Optim. Online*, 2013.

[8] P. M. Esfahani and D. Kuhn, Data-driven distributionally robust optimization using the wassersttein metric: performance guarantees and tractable reformulations, *Optim. Online*, 2015.

[9] K. Fan, Minimax theorems, *Proceedings of National Academy of Sciences of the United States of America*, Vol. 39, pp. 42-47, 1953.

[10] A. L. Gibbs and F. E. Su, On choosing and bounding probability metrics, *International Statistical Review*, Vol. 70, pp. 419-435,

[11] D. Klatte, A note on quantitative stability results in nonlinear optimization, *Seminarbericht Nr. 90*, Sektion Mathematik, Humboldt-Universität zu Berlin, Berlin, pp. 77-86, 1987.

[12] J. A. Mirrlees, The theory of moral hazard and unobservable behaviour: Part I, *Review of Economic Studies*, Vol. 66, pp. 3-21, 1999.

[13] S. Mehrotra and D. Papp, A cutting surface algorithm for semiinfinite convex programming with an application to moment robust optimization, *SIAM J. Optim.*, 24: 1670-1697, 2014.

[14] G. Ch. Pflug and A. Pichler, Approximations for probability distributions and stochastic optimization problems, M. Bertocchi, G. Consigli and M. A. H. Dempster, eds. *Stochastic Optimization Methods in Finance and Energy*, Vol. 163, Springer, Berlin, 2011

[15] G. Ch. Pflug and D. Wozabal. Ambiguity in portfolio selection. *Quant. Financ.*, 7: 435-442, 2007.

[16] M. Riis and K. A. Andersen, Applying the minimax criterion in stochastic recourse programs, *European Journal of Operational Research*, Vol. 165, pp. 569-584, 2005.

[17] R. T. Rockafellar and R. J-B. Wets, *Variational analysis*, Springer, Berlin, 1998.

[18] W. Römisch, Stability of stochastic programming problems, in *Stochastic Programming*, A. Ruszczyński and A. Shapiro, eds., Elsevier, Amsterdam, pp. 483-554, 2003.

[19] A. Shapiro, On duality theory of conic linear problems, Miguel A. Goberna and Marco A. López, eds., *Semi-Infinite Programming: Recent Advances, Kluwer Academic Publishers*, pp. 135-165, 2001.

[20] A. Shapiro, D. Dentcheva and A. Ruszczyński, *Lectures on Stochastic Programming: Modeling and Theory*, SIAM, Philadelphia, PA, 2009.

[21] H. Sun and H. Xu, Convergence analysis for distributionally robust optimization and equilibrium problems, *Mathematics of Operations Research*, 2015.

[22] S. Takriti, S. Ahmed, Managing short-term electricity contracts under uncertainty: A minimax approach. $http://www.isye.gatech.edu/\ sahmed$

[23] H. Xu, Y. Liu and H. Sun, Distributionally robust optimization with matrix moment constraints: Lagrange duality and cutting plane methods, Optimization Online, 2015.

[24] J. Zhang, H. Xu and L. Zhang, Quantitative stability analysis of stochastic quasi-variational inequality problems and applications, Optimization Online, 2015.

[25] C. Zhao and Y. Guan, Data-driven risk-averse stochastic optimization with wasserstein metric, Optimization Online, 2015.

6 Appendix

Recall that for a nonempty set $S \subseteq \mathbb{R}^n$, the *horizon cone* of S is given by

$$S^\infty := \{x : \exists x_k \in C, \lambda_k \searrow 0, \text{with } \lambda_k x_k \searrow x\}.$$

For function $f : \mathbb{R}^n \to \mathbb{R} \cup \pm\infty$, the associated *horizon function* $f^\infty : \mathbb{R}^n \to \mathbb{R} \cup \pm\infty$ is a function whose epigraph coincides with the horizon cone of the epigraph of f, i.e.,

$$\text{epi } f^\infty := (\text{epi } f)^\infty \text{ if } f \not\equiv +\infty$$

and $f^\infty = \delta_{\{0\}}$ if $f \equiv +\infty$, where

$$\delta_A(x) := \begin{cases} 0 & \text{if } x \in A, \\ +\infty & \text{otherwise.} \end{cases}$$

See [17, Definitions 3.3 and 3.17].

Lemma 6.1 *([17, Theorem 3.21]) For $f : \mathbb{R}^n \to \mathbb{R} \cup \pm\infty$, if $f \not\equiv +\infty$, then*

$$f^\infty(w) = \lim_{\delta \searrow 0} \inf_{\lambda \in (0,\delta), x \in B(w,\delta)} \lambda f(\lambda^{-1} x).$$

Lemma 6.2 *([17, Exercise 3.29]) Let h_1 and h_2 be lower semicontinuous (lsc for short) and proper on $\mathbb{R}^n$, and suppose $h_1^\infty(0) \neq -\infty$ and $h_2^\infty(0) \neq -\infty$. Then*

$$(h_1 + h_2)^\infty \geq h_1^\infty + h_2^\infty.$$

Definition 6.1 *([17, Definition 1.16]) A function $h : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \cup \pm\infty$ with values $h(x, u)$ being level-bounded in x locally uniformly in u if for each $\bar{u} \in \mathbb{R}^m$ and $\alpha \in \mathbb{R}$, there is a neighborhood V of $\bar{u}$ along with a bounded set $B \subset \mathbb{R}^n$ such that $\{x : h(x, u) \leq \alpha\} \subset B$ for all $u \in V$.*

Lemma 6.3 *([17, Theorem 1.17]) Consider*

$$p(u) := \inf_x h(x, u), \qquad P(u) := \arg\min_x h(x, u),$$

where $h : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \cup \pm\infty$ is a proper, lsc function. If (a) $h(x, u)$ is level-bounded in x locally uniformly in u, (b) there exists some $\bar{x} \in P(\bar{u})$ such that $h(\bar{x}, u)$ is continuous in u at $\bar{u}$ relative to a set U cintianing $\bar{u}$, then $p(\cdot)$ is continuous at $\bar{u}$.

Lemma 6.4 *([17, Theorem 3.31]) For a proper, lsc function $h : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \cup \pm\infty$, a sufficient condition for $h(x, u)$ to be level-bounded in x locally uniformly in u, which is also necessary when h is convex, is*

$$h^\infty(x, 0) > 0 \quad \text{for all } x \neq 0.$$

Let $\phi(x,u,\Lambda,\xi) := -f(x,\xi) - \Lambda \bullet \Psi(x,u,\xi)$. We investigate continuity of ψ by virtue of Lemmas 6.1-6.4 .

Proposition 6.1 *(**Continuity of** ψ) Let $(x_0,u_0) \in X \times U$ be fixed. Assume ϕ is a proper and lsc function on $X \times U \times \mathbb{R}^k \times \mathcal{K}$. Suppose (a) $\Psi(\cdot,\cdot,\xi)$ and $f(\cdot,\xi)$ are continuous at (x_0,u_0) and x_0 respectively for each $\xi \in \Xi$, (b)*

$$\phi^\infty(0,0,0,0) \neq -\infty \text{ and } \phi^\infty(0,0,0,\xi) > 0 \text{ for all } \xi \in \Xi^\infty\backslash\{0\}. \tag{6.87}$$

Then ψ is continuous at (x_0,u_0,Λ_0) for all $\Lambda_0 \in K^*$.

Proof. Observe first that

$$\psi(x,u,\Lambda,\xi) = -\inf_{\xi\in\mathbb{R}^k}[\phi(x,u,\Lambda,\xi) + \delta_\Xi(\xi)].$$

We only need to verify the conditions in Lemma 6.3. Let $H(x,u,\Lambda,\xi) := \phi(x,u,\Lambda,\xi) + \delta_\Xi(\xi)$. Then H is a proper, lsc function. By Lemma 6.1, $\delta_\Xi^\infty(\cdot) = \delta_{\Xi^\infty}(\cdot)$. Since $\phi^\infty(0,0,0,0) \neq -\infty$ by assumption, it follows by Lemma 6.2 that

$$H^\infty(x,u,\Lambda,\cdot) \geq \phi^\infty(x,u,\Lambda,\cdot) + \delta_\Xi^\infty(\cdot) = \phi^\infty(x,u,\Lambda,\cdot) + \delta_{\Xi^\infty}(\cdot),$$

which, by condition (b), implies

$$H^\infty(x,u,\Lambda,\xi) > 0, \ \forall \xi \in \Xi^\infty\backslash\{0\}.$$

By Lemma 6.4, the latter ensures that H is level-bounded in ξ locally uniformly in (x,u,Λ), hence condition (a) of Lemma 6.3 is fulfilled. On the other hand, under condition (a), $H(x,u,\Lambda,\xi)$ is continuous at (x_0,u_0,Λ_0) for every $\Lambda_0 \in \mathcal{K}^*$ and each $\xi \in \Xi$, which verifies condition (b) in Lemma 6.3. ■

With Proposition 6.1, we can derive the following sufficient conditions for the function $f(x,\cdot) + \Lambda \bullet \Psi(x,u,\cdot)$ achieves its minimum in a compact subset of Ξ uniformly w.r.t. (x,u,Λ).

Proposition 6.2 *Suppose conditions in Proposition 6.1 holds. If*

$$[f(x_0,\cdot) + \Lambda \bullet \Psi(x_0,u_0,\cdot)]^\infty(\xi) > 0, \quad \forall \xi \in \Xi^\infty\backslash\{0\}, \quad \forall \Lambda \in \mathcal{K}^*, \tag{6.88}$$

then there exist a compact set $\hat{\Xi} \subset \Xi$ such that

$$\psi(x,u,\Lambda) = \sup_{\xi\in\hat{\Xi}}[f(x,\xi) + \Lambda \bullet \Psi(x,u,\xi)]$$

for (x,u) close to (x_0,u_0) and $\Lambda \in \mathcal{K}^*$.

Proof. Assume for the sake of a contradiction that this is not true. Then there exist a $\Lambda_0 \in \mathcal{K}^*$ and $(x_k,u_k) \to (x_0,u_0), \xi_k \in \Xi$ such that $\|\xi_k\| \to +\infty$ and

$$\psi(x_k,u_k,\Lambda_0) \geq f(x_k,\xi_k) + \Lambda_0 \bullet \Psi(x_k,u_k,\xi_k). \tag{6.89}$$

Without loss of generality, we assume $\lim_{k\to\infty} \xi_k/\|\xi_k\| = \hat{\xi}$ and let $\lambda_k = \frac{1}{\|\xi_k\|}$. Then $\hat{\xi} \in \Xi^\infty\backslash\{0\}$ and $\lambda_k \to 0$. Multiplying λ_k both sides of inequality (6.89), we obtain

$$\lambda_k\psi(x_k,u_k,\Lambda_0) \geq \lambda_k f\left(x_k, \lambda_k^{-1}\frac{\xi_k}{\|\xi_k\|}\right) + \lambda_k\Lambda_0 \bullet \Psi\left(x_k,u_k,\lambda_k^{-1}\frac{\xi_k}{\|\xi_k\|}\right).$$

Note that under condition (6.88), it follows by Proposition 6.1 that $\psi(x_k, u_k, \Lambda_0)$ is bounded. By driving k to infinity, the inequality above yields

$$\begin{aligned} 0 &\geq \liminf_{k\to\infty} \lambda_k \psi(x_k, u_k, \Lambda_0) \geq \liminf_{k\to\infty} \lambda_k f\left(x_k, \lambda_k^{-1}\frac{\xi_k}{\|\xi_k\|}\right) + \lambda_k \Lambda_0 \bullet \Psi\left(x_k, u_k, \lambda_k^{-1}\frac{\xi_k}{\|\xi_k\|}\right) \\ &\geq [f(x_0,\cdot)] + \Lambda \bullet \Psi(x_0, u_0, \cdot)]^\infty(\hat{\xi}), \end{aligned}$$

which means that $\hat{\xi} = 0$, a contradiction! ■

Condition (6.88) is a key condition in Proposition 6.2, it is therefore might be helpful to discuss how the condtion may be fultiled. First, by Lemma 6.2, if $f(x_0,\cdot)^\infty(0) \neq -\infty$, $[\Lambda \bullet \Psi(x_0, u_0, \cdot)]^\infty(0) \neq -\infty$ and

$$f(x_0,\cdot)^\infty(\xi) + [\Lambda \bullet \Psi(x_0, u_0, \cdot)]^\infty(\xi) > 0, \quad \forall \xi \in \Xi^\infty \backslash \{0\}, \quad \forall \Lambda \in \mathcal{K}^*,$$

then condition (6.88) is satisfied. Second, if $f(x_0,\cdot)$ and $\Psi(x_0,u_0,\cdot)$ are continuous, $f(x_0,\cdot) + \Lambda \bullet \Psi(x_0,u_0,\cdot)$ is positively homogeneous, i.e., $0 \in \text{dom } [f(x_0,\cdot) + \Lambda \bullet \Psi(x_0,u_0,\cdot)]$ and $f(x_0,a\xi) + \Lambda \bullet \Psi(x_0,u_0,a\xi) = a[f(x_0,\xi) + \Lambda \bullet \Psi(x_0,u_0,\xi)]$ for all ξ and constant $a > 0$,

$$f(x_0,\xi) + \Lambda \bullet \Psi(x_0, u_0, \xi) > 0, \quad \forall \xi \in \Xi^\infty \backslash \{0\}, \quad \forall \Lambda \in \mathcal{K}^*,$$

then condition (6.88) is satisfied. Indeed, by the discussion after [17, Exercise 3.19], $[f(x_0,\cdot) + \Lambda \bullet \Psi(x_0,u_0,\cdot)]^\infty = f(x_0,\cdot) + \Lambda \bullet \Psi(x_0,u_0,\cdot)$ in this case. Third, to see a concrete example for the condition, we may consider a simple example where $f(x_0,\xi) = b^T\xi$, $\Psi(x_0,u_0,\xi) = \frac{1}{2}\xi\xi^T$ and $\mathcal{K} = S_+^{k\times k}$ with $b \in \mathbb{R}^k, \xi \in \mathbb{R}^k$. If $\vartheta(\Lambda) > 0$ for all $\Lambda \in S_+^{k\times k}$, then condition (6.88) holds, where $\vartheta(\Lambda)$ is the optimal value of the following parametric problem:

$$\begin{aligned} \min_{\xi} \quad & b^T\xi \\ \text{s.t.} \quad & \Lambda\xi = 0, \\ & \xi \in \Xi^\infty \backslash \{0\}. \end{aligned}$$

Note that in this case, we deduce from a disscussion on horizon properties for quadratic functions after [17, Theorem 3.21],

$$[f(x_0,\cdot) + \Lambda \bullet \Psi(x_0,u_0,\cdot)]^\infty(\xi) = b^T\xi + \delta_{\{\xi:\Lambda\xi=0\}}(\xi).$$