

## **Optimality conditions and Moreau–Yosida regularization for almost sure state constraints**

Caroline Geiersbach<sup>1</sup>, Michael Hintermüller<sup>1, 2</sup>

submitted: August 3, 2021

<sup>1</sup> Weierstrass Institute  
Mohrenstr. 39  
10117 Berlin  
Germany

E-Mail: [caroline.geiersbach@wias-berlin.de](mailto:caroline.geiersbach@wias-berlin.de)  
[michael.hintermueller@wias-berlin.de](mailto:michael.hintermueller@wias-berlin.de)

<sup>2</sup> Humboldt-Universität zu Berlin  
Unter den Linden 6  
10099 Berlin  
Germany

E-Mail: [hint@math.hu-berlin.de](mailto:hint@math.hu-berlin.de)

No. 2862

Berlin 2021



---

2020 *Mathematics Subject Classification.* 49K20, 49K45, 49N15, 49J20, 90C15.

*Key words and phrases.* PDE-constrained optimization under uncertainty, optimization in Banach spaces, optimality conditions, regularization, convex stochastic optimization in Banach spaces, two-stage stochastic optimization, duality.

Edited by  
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)  
Leibniz-Institut im Forschungsverbund Berlin e. V.  
Mohrenstraße 39  
10117 Berlin  
Germany

Fax: +49 30 20372-303  
E-Mail: [preprint@wias-berlin.de](mailto:preprint@wias-berlin.de)  
World Wide Web: <http://www.wias-berlin.de/>

# Optimality conditions and Moreau–Yosida regularization for almost sure state constraints

Caroline Geiersbach, Michael Hintermüller

## Abstract

We analyze a potentially risk-averse convex stochastic optimization problem, where the control is deterministic and the state is a Banach-valued essentially bounded random variable. We obtain strong forms of necessary and sufficient optimality conditions for problems subject to equality and conical constraints. We propose a Moreau–Yosida regularization for the conical constraint and show consistency of the optimality conditions for the regularized problem as the regularization parameter is taken to infinity.

## 1 Introduction

Let  $X_1$  and  $X_2$  be real, reflexive, and separable Banach spaces, and let  $(\Omega, \mathcal{F}, \mathbb{P})$  denote a probability space, where  $\Omega$  represents the sample space,  $\mathcal{F} \subset 2^\Omega$  is the  $\sigma$ -algebra of events on the power set of  $\Omega$ , and  $\mathbb{P}: \Omega \rightarrow [0, 1]$  is a probability measure. Consider a convex stochastic optimization problem of the form

$$\begin{aligned} & \underset{x_1, x_2(\cdot)}{\text{minimize}} && J_1(x_1) + \mathcal{R}[J_2(x_1, x_2(\cdot); \cdot)] \\ & \text{subject to (s.t.)} && \begin{cases} x_1 \in C, \\ x_2(\omega) \in X_{2,\text{ad}}(x_1, \omega) \quad \text{a.s.}, \end{cases} \end{aligned} \tag{1}$$

where  $C \subset X_1$  is nonempty, closed, and convex;  $X_{2,\text{ad}}(x_1, \omega) \subset X_2$  is assumed to be nonempty, closed, and convex for all  $x_1 \in C$  and almost all  $\omega \in \Omega$ . In the problem (1), the control  $x_1$  is deterministic and the state  $x_2$  is a vector-valued random variable. This is the natural paradigm if one views  $x_1$  as a concrete “decision” that should be made while taking into account uncertainties from inputs or parameters in the state system. In the classical stochastic optimization nomenclature,  $x_1$  and  $x_2$  are called first and second stage variables, respectively; see, e.g., [31, 41] for an introduction to two-stage stochastic optimization.

There are numerous applications ranging from economics and finance [29, 30], engineering [14, 35], and life sciences [15] that can be mathematically expressed as (1). Often, historical information (in the form of a probability distribution) about uncertain demand, performance, or physical parameters has been collected and one would like to take this information into account while making a decision. In some settings, it is enough to optimize over these possibilities on average, in which case one works with a risk-neutral formulation, meaning the measure  $\mathcal{R}$  is simply the expectation, or average. However, there are many applications where the average is not desirable, and one works instead with a different choice of risk functional  $\mathcal{R}: L^p(\Omega) \rightarrow \mathbb{R} \cup \{\infty\}$ , which captures the optimizer’s desired risk tolerance.

For many applications, the spaces  $X_1$  and  $X_2$  are Euclidean spaces, which simplifies the analysis considerably. For instance, in portfolio optimization, the decision variable may be a (finite) vector of

real-valued weights representing the percentages of total cash dedicated to each asset. The second-stage variable is another vector corresponding to a redistribution of assets after new information is gathered. However, there are some problems in which the underlying spaces are infinite-dimensional, for instance, in optimization with partial differential equations (PDEs) subject to uncertainty; see e.g., [2, 8, 13, 18, 24, 25, 27, 40], and the references therein. As an example, consider the optimal control of a stationary heat source over a domain  $D \subset \mathbb{R}^d$ , where the conductivity  $a = a(s, \omega)$  is uncertain, but known to follow a given probability distribution. One seeks a control  $x_1$  and the resulting solution  $x_2 = x_2(s, \omega)$  of the PDE, the latter of which does not exceed a threshold  $\psi$  and, on average, best approximates some target distribution  $y_D$ . With  $\alpha \geq 0$ , the model is formally expressed as

$$\begin{aligned} & \underset{x_1, x_2(\cdot)}{\text{minimize}} && \frac{1}{2} \mathbb{E} \left[ \|x_2 - y_D\|_{L^2(D)}^2 \right] + \frac{\alpha}{2} \|x_1\|_{L^2(D)}^2 \\ & \text{s.t.} && \begin{cases} x_1 \in C, \\ -\nabla \cdot (a(s, \omega) \nabla x_2(s, \omega)) = x_1(s) & \text{on } D \times \Omega, \\ x_2(s, \omega) = 0 & \text{on } \partial D \times \Omega, \\ x_2(s, \omega) \leq \psi(s, \omega) & \text{on } D \times \Omega, \end{cases} \end{aligned} \quad (\text{P}')$$

where the constraints are to hold almost everywhere in  $D$  and almost surely in  $\Omega$ . Other examples of conical constraints will be presented in Section 4.1. To develop optimality conditions for this problem, a constraint qualification is needed, which implicitly requires that subsets of the given function space have a nonempty interior. Standard PDE theory (cf. [1]) gives, under mild assumptions on  $a$  and  $D$ , a unique solution  $x_2(\cdot, \omega) \in X_2 := H_0^1(D) \cap H^2(D)$  for every  $x_1 \in X_1 := L^2(D)$  and  $\omega \in \Omega$ . Through embeddings one obtains even  $x_2(\cdot, \omega) \in C(\bar{D})$ , where an interior point condition is satisfiable. However, without additional structure on  $\Omega$ ,  $x_2$  as a function in  $L^p(\Omega, X_2)$  can only satisfy an interior point condition if  $p = \infty$ . It is notable that for a deterministic model, i.e., where  $x_2 = x_2(s)$ , the Lagrange multiplier associated with the conical constraint is generally only a measure. It is standard to employ regularization in numerical procedures to circumvent this difficulty; see [21, 22].

In this paper, we are concerned with obtaining strong forms of optimality conditions for a problem like (1), by which we mean explicit, “almost sure”-type conditions, which are amenable to building algorithms. The main difficulty here is in applying classical duality theory, where one would expect a Lagrange multiplier to belong to the space  $(L^\infty(\Omega, X))^*$ ,  $X$  being a Banach space. A multiplier from this space cannot generally be represented as a mapping from  $\Omega$  to, say,  $X^*$ . Hence the task of establishing optimality conditions in the strong form, i.e., with almost sure conditions, is quite challenging. Another difficulty is in the structure of the problem itself, where the optimality conditions should make explicit the first stage variable’s independence of  $\omega$ . Luckily, insights can be found in classical literature, above all in the important works by Rockafellar and Wets [36, 37, 38, 39], who developed theory for the case  $X = \mathbb{R}^d$ . Recently, this theory was revisited in [19] with a focus on the more general vector-valued case. The authors proved the existence of integrable Lagrange multipliers under a relatively complete recourse condition (defined in Assumption 3.6). For many applications, including the problem (P’), this assumption is too strong. In [19], a regularization was proposed and consistency of the primal problem was analyzed as the penalization parameter is taken to infinity. However, a systematic analysis for a more general problem class was missing.

Optimality conditions with an application to PDE-constrained optimization under uncertainty were first presented in [26]. While the framework in [26] allows for generally nonconvex problems, the case where the feasible set  $X_{2,\text{ad}}(\omega)$  contains conical constraints is not covered. Our work is related to a recent preprint by [17]. As part of this work, a risk-neutral model problem from optimal control is presented and the authors also propose a Moreau–Yosida regularization to solve their problem. To this end, they impose more structure on their probability space, which allows for standard arguments via

embeddings. However, they do not obtain strong forms of optimality conditions, which can yield deep insight into the structure of the optimality conditions for the original problem. The framework presented in our work is general enough to cover the theory in [17] without additional structure on the probability space. It is worth mentioning that our work related to the recent trend to include chance constraints in PDE-constrained optimization under uncertainty [12, 16, 20]. These models are of interest when deviations from a hard constraint are permissible with a certain probability. Notably, the chance constraint and regularized “almost sure” models can be shown to be related to one another [17].

We proceed as follows. In Section 2, we present notation and technical results needed for the handling of essentially bounded Banach-valued random variables. Additionally, we recall key properties of risk functionals. In Section 3, we generalize the risk-neutral framework given in [19] by allowing for potentially risk-averse objective functions. Additionally, we present necessary and sufficient conditions without the assumption of relatively complete recourse. While regular (integrable) Lagrange multipliers do not exist in this case, we show that Lagrange multipliers can be split into regular parts and singular terms, which can be handled separately. In Section 4, we present the Moreau–Yosida regularization of almost sure conical constraints and show that the optimality system for the regularized problem is consistent with the original problem as the penalization parameter is taken to infinity. An outlook is given in Section 5.

## 2 Preliminaries

### 2.1 Background and Notation

Throughout the paper, we denote the dual space of a (real) Banach space  $X$  by  $X^*$  and the canonical dual pairing by  $\langle \cdot, \cdot \rangle_{X^*, X}$ . If  $X$  is paired with another space  $U$ , we write  $\langle \cdot, \cdot \rangle_{U, X}$  to denote their pairing. The symbols  $\rightarrow$ ,  $\rightharpoonup$ , and  $\rightharpoonup^*$  denote strong, weak, and weak\* convergence, respectively;  $A : X \rightrightarrows Y$  denotes a set-valued operator from  $X$  to  $Y$ . We recall that for a proper function  $h : X \rightarrow \mathbb{R} \cup \{\infty\}$ , the subdifferential (in the sense of convex analysis) is the set-valued operator

$$\partial h : X \rightrightarrows X^* : x \mapsto \{q \in X^* : \langle q, y - x \rangle_{X^*, X} + h(x) \leq h(y) \quad \forall y \in X\}.$$

The domain and epigraph of  $h$  are denoted by  $\text{dom}(h) = \{x \in X : h(x) < \infty\}$  and  $\text{epi}(h) = \{(x, \alpha) \in X \times \mathbb{R} : h(x) \leq \alpha\}$ , respectively. The sum of two sets  $A$  and  $B$  with  $\lambda \in \mathbb{R}$  is given by  $A + \lambda B := \{a + \lambda b : a \in A, b \in B\}$ . A nonempty subset  $K$  of  $X$  is said to be a cone if for any  $x \in K$ ,  $t \geq 0$ , it follows that  $tx \in K$ . The indicator function of a set  $C$  is denoted by  $\delta_C$  and is defined by  $\delta_C(x) = 0$  if  $x \in C$  and  $\delta_C(x) = \infty$  otherwise. For a nonempty and convex set  $C \subset X$ , the normal cone  $N_C(x)$  at  $x \in C$  is defined by  $N_C(x) := \{z \in X^* : \langle z, y - x \rangle_{X^*, X} \leq 0 \quad \forall y \in C\}$ . We set  $N_C(x) := \emptyset$  if  $x \notin C$ . Note that  $\partial \delta_C(x) = N_C(x)$  for all  $x \in C$ .

Given a Banach space  $X$  equipped with the norm  $\|\cdot\|_X$ , the Bochner space  $L^r(\Omega, X) := L^r(\Omega, \mathcal{F}, \mathbb{P}; X)$  is the set of all (equivalence classes of) strongly measurable functions  $y : \Omega \rightarrow X$  having finite norm, where the norm is given by

$$\|y\|_{L^r(\Omega, X)} := \begin{cases} (\int_{\Omega} \|y(\omega)\|_X^r \, d\mathbb{P}(\omega))^{1/r}, & r < \infty \\ \text{ess sup}_{\omega \in \Omega} \|y(\omega)\|_X, & r = \infty \end{cases}.$$

If  $X = \mathbb{R}$  we use the shorthand  $L^r(\Omega)$ . An  $X$ -valued random variable  $x$  is Bochner integrable if there exists a sequence  $\{x_n\}$  of  $\mathbb{P}$ -simple functions  $x_n : \Omega \rightarrow X$  such that  $\lim_{n \rightarrow \infty} \int_{\Omega} \|x_n(\omega) -$

$x(\omega)\|_X \, d\mathbb{P}(\omega) = 0$ . The limit of the integrals of  $x_n$  gives the Bochner integral (the expectation), i.e.,

$$\mathbb{E}[x] := \int_{\Omega} x(\omega) \, d\mathbb{P}(\omega) = \lim_{n \rightarrow \infty} \int_{\Omega} x_n(\omega) \, d\mathbb{P}(\omega).$$

Clearly, this expectation is an element of  $X$ .

Throughout the paper, we work with parametrized functions of the type  $f : X \times \Omega \rightarrow Y$ , and write  $f(x; \omega)$  for the mapping from  $X$  to  $Y$  for a fixed parameter  $\omega \in \Omega$ . A property is said to hold almost surely (a.s.) provided that the set (in  $\Omega$ ) where the property does not hold is a null set. As an example, two random variables  $\xi, \xi'$  are said to be equal almost surely,  $\xi = \xi'$  a.s., if and only if  $\mathbb{P}(\{\omega \in \Omega : \xi(\omega) \neq \xi'(\omega)\}) = 0$ . Sometimes we write “for almost every  $\omega$ ” to mean a.s.

## 2.2 Properties of $L^\infty(\Omega, X)$

In this section, we present some technical results that will be crucial to characterizing optimality for problems like (1). Throughout this section, let  $X$  be an arbitrary (real) reflexive and separable Banach space, and let the  $\sigma$ -algebra of Borel sets on  $X$  be denoted by  $\mathcal{B}$ . A function  $f : X \times \Omega \rightarrow \mathbb{R}$  is called a convex integrand if  $f(\cdot; \omega)$  is convex for almost every  $\omega$ . Following Rockafellar [33], this integrand is called *normal* if it is not identically infinity, it is  $(\mathcal{B} \times \mathcal{F})$ -measurable, and  $f(\cdot; \omega)$  is lower semicontinuous in  $X$  for each  $\omega \in \Omega$ . An example of a function that is normal is one that is finite everywhere and Carathéodory, meaning  $f$  is measurable in  $\omega$  for fixed  $x$  and continuous in  $x$  for fixed  $\omega$ . Normality of  $f$  implies that it is superpositionally measurable, meaning  $\omega \mapsto f(x(\omega); \omega)$  is measurable as long as  $x : \Omega \rightarrow X$  is measurable.

For a normal convex integrand  $f$ , we define the integral functional on  $L^\infty(\Omega, X)$  by

$$I_f(x) = \int_{\Omega} f(x(\omega); \omega) \, d\mathbb{P}(\omega),$$

i.e., the expectation of the superposition of  $f$ . A continuous linear functional  $v \in (L^\infty(\Omega, X))^*$  of the form

$$\langle v, x \rangle_{(L^\infty(\Omega, X))^*, L^\infty(\Omega, X)} = \int_{\Omega} \langle x^*(\omega), x(\omega) \rangle_{X^*, X} \, d\mathbb{P}(\omega) \quad (2)$$

for some  $x^* \in L^1(\Omega, X^*)$  is said to be absolutely continuous. A proof of the following result is in [19].

**Lemma 2.1.** Assume  $f$  is a normal convex integrand and  $f(x(\omega); \omega)$  is an integrable function of  $\omega$  for every  $x \in L^\infty(\Omega, X)$ . Then  $I_f$  and  $I_{f^*}$  are well-defined convex functionals on  $L^\infty(\Omega, X)$  and  $L^1(\Omega, X^*)$ , respectively, that are conjugate to each other in the sense that

$$I_{f^*}(x^*) = \sup_{x \in L^\infty(\Omega, X)} \left\{ \int_{\Omega} \langle x^*(\omega), x(\omega) \rangle_{X^*, X} \, d\mathbb{P}(\omega) - I_f(x) \right\},$$

$$I_f(x) = \sup_{x^* \in L^1(\Omega, X^*)} \left\{ \int_{\Omega} \langle x^*(\omega), x(\omega) \rangle_{X^*, X} \, d\mathbb{P}(\omega) - I_{f^*}(x^*) \right\}.$$

Furthermore, if  $v^*$  is an absolutely continuous functional corresponding to a function  $x^* \in L^1(\Omega, X^*)$ , then  $I_f^*(v^*) = I_{f^*}(x^*)$ , while  $I_f^*(v^*) = \infty$  for any  $v^*$  that is not absolutely continuous.

**Remark 2.2.** Since  $X$  is reflexive, it satisfies the Radon–Nikodym property; moreover, the underlying probability space is  $\sigma$ -finite. It follows that the dual space  $(L^1(\Omega, X^*))^*$  can be identified with  $L^\infty(\Omega, X)$ . Thus the mapping

$$x^* \mapsto \int_{\Omega} \langle x^*(\omega), x(\omega) \rangle_{X^*, X} \, d\mathbb{P}(\omega) = \langle x^*, x \rangle_{L^1(\Omega, X^*), L^\infty(\Omega, X)}$$

is weakly\* continuous and convex. Moreover,  $I_{f^*}$ , as the supremum over  $(x, \alpha) \in \text{epi}(I_f)$  of functionals of the form

$$x^* \mapsto \langle x^*, x \rangle_{L^1(\Omega, X^*), L^\infty(\Omega, X)} - \alpha,$$

is convex and weakly\* lower semicontinuous. A consequence of Lemma 2.1 is that in the case where  $f$  is a normal convex integrand and  $f(x(\omega); \omega)$  is integrable for all  $x \in L^\infty(\Omega, X)$ , we have  $I_f^{**} \equiv I_{f^*} \equiv I_f$  and thus  $I_f$  is also weakly\* lower semicontinuous. Another helpful method to prove weak\* lower semicontinuity of  $I_f$  follows from [33, p. 227]: if  $f$  is a normal convex integrand and neither  $I_f$  nor  $I_{f^*}$  are identically infinity, then  $I_f$  and  $I_{f^*}$  are conjugate to each other.

In this work, we make extensive use of a decomposition of elements belonging to  $(L^\infty(\Omega, X))^*$ . Note first that the space of absolutely continuous functionals defined by (2) form a closed subspace of  $(L^\infty(\Omega, X))^*$  that is isometric to  $L^1(\Omega, X^*)$ . This subspace has a complement consisting of singular functionals, given in the next definition.

**Definition 2.3.** A functional  $v^\circ \in (L^\infty(\Omega, X))^*$  is called *singular (relative to  $\mathbb{P}$ )* if there exists a sequence  $\{F_n\} \subset \mathcal{F}$  with  $F_{n+1} \subset F_n$  for all  $n$  and  $\mathbb{P}(F_n) \rightarrow 0$  as  $n \rightarrow \infty$  such that  $v^\circ$  is concentrated on  $\{F_n\}$ , i.e.,  $\langle v^\circ, x \rangle_{(L^\infty(\Omega, X))^*, L^\infty(\Omega, X)} = 0$  for all  $x \in L^\infty(\Omega, X)$  that vanishes on some  $F_n$ .

The following Yosida–Hewitt-type decomposition result for vector measures was proven in [23, Appendix 1, Theorem 3] (with a slight correction to the original proof in [28]). A related result is in [42, Theorem 4].

**Theorem 2.4** (Ioffe and Levin). Each functional  $v^* \in (L^\infty(\Omega, X))^*$  has a unique decomposition

$$v^* = v + v^\circ,$$

where  $v$  is absolutely continuous,  $v^\circ$  is singular relative to  $\mathbb{P}$ , and

$$\|v^*\|_{(L^\infty(\Omega, X))^*} = \|v\|_{(L^\infty(\Omega, X))^*} + \|v^\circ\|_{(L^\infty(\Omega, X))^*}.$$

Using this decomposition, it is possible to obtain a particularly useful expression for the subdifferential of integral functionals on  $L^\infty(\Omega, X)$ . The next result was proven in [19, Corollary 2.7].

**Lemma 2.5.** Suppose  $f$  is a normal convex integrand. Let  $\bar{x} \in L^\infty(\Omega, X)$  be a point at which there exists  $r > 0$  and  $k_r \in L^1(\Omega)$  satisfying  $f(x(\omega); \omega) \leq k_r(\omega)$  as long as  $\|x - \bar{x}\|_{L^\infty(\Omega, X)} < r$ . Then  $v^* \in (L^\infty(\Omega, X))^*$  is an element of  $\partial I_f(\bar{x})$  if and only if

$$x^*(\omega) \in \partial f(\bar{x}(\omega); \omega) \quad \text{a.s.}, \tag{3}$$

where  $x^* \in L^1(\Omega, X^*)$  corresponds to the absolutely continuous part  $v$  of  $v^*$ . Moreover,  $\partial I_f(\bar{x})$  can be identified with a nonempty, weakly compact subset of  $L^1(\Omega, X^*)$ . In particular,  $v^*$  belongs to  $\partial I_f(\bar{x})$  if and only if  $v^\circ \equiv 0$  and  $v \equiv x^*$  satisfies (3).

To close this section, we present some auxiliary results for operator-valued random variables that will be of use later in the paper. Let  $\mathcal{L}(Y, Z)$  denote the space of all bounded linear operators from  $Y$  to  $Z$  (both real, reflexive, and separable Banach spaces). The set  $L^\infty(\Omega, \mathcal{L}(Y, Z))$  corresponds to all strongly measurable operator-valued random variables  $\mathcal{A} : \Omega \rightarrow \mathcal{L}(Y, Z)$  such that

ess sup $_{\omega \in \Omega} \|\mathcal{A}(\omega)\|_{\mathcal{L}(Y, Z)} < \infty$ . The adjoint and inverse operators are to be understood in the “almost sure” sense; e.g., for  $\mathcal{A}$ , the adjoint operator is the (strongly measurable) random operator  $\mathcal{A}^*$  such that for all  $(y, z^*) \in Y \times Z^*$ ,

$$\mathbb{P}(\{\omega \in \Omega : \langle z^*, \mathcal{A}(\omega)y \rangle_{Z^*, Z} = \langle \mathcal{A}^*(\omega)z^*, y \rangle_{Y^*, Y}\}) = 1.$$

The next two results are proven in Appendix A. In the proofs, and in other arguments throughout the paper, we make use of the fact that bounded sets in  $L^\infty(\Omega, Y)$  (the dual of a separable space) equipped with the weak\* topology are metrizable; the same applies for  $L^\infty(\Omega, Z)$ . In particular, topological and sequential notions coincide.

**Lemma 2.6.** Let  $\mathcal{A} \in L^\infty(\Omega, \mathcal{L}(Y, Z))$ . Then

- (i) The operators  $\mathcal{A}_1: L^\infty(\Omega, Y) \rightarrow L^\infty(\Omega, Z), y(\cdot) \mapsto \mathcal{A}(\cdot)y(\cdot)$  and  $\mathcal{A}_1^*: (L^\infty(\Omega, Z))^* \rightarrow (L^\infty(\Omega, Y))^*$  are bounded. Additionally,  $\mathcal{A}_1^*$  maps singular elements of  $(L^\infty(\Omega, Z))^*$  to singular elements of  $(L^\infty(\Omega, Y))^*$ . Moreover,  $\hat{\mathcal{A}}_1^*: L^1(\Omega, Z^*) \rightarrow L^1(\Omega, Y^*), z^*(\cdot) \mapsto \mathcal{A}^*(\cdot)z^*(\cdot)$  is bounded.
- (ii) The operators  $\mathcal{A}_2: Y \rightarrow L^\infty(\Omega, Z), y \mapsto \mathcal{A}(\cdot)y$  and  $\mathcal{A}_2^*: (L^\infty(\Omega, Z))^* \rightarrow Y^*$  are bounded. Moreover,  $\hat{\mathcal{A}}_2^*: L^1(\Omega, Z^*) \rightarrow L^1(\Omega, Y^*), z^*(\cdot) \mapsto \mathcal{A}^*(\cdot)z^*(\cdot)$  is bounded.

In Lemma 2.6, the notation  $\mathcal{A}(\cdot)y(\cdot)$  signifies an element of  $L^\infty(\Omega, Z)$  with the pointwise action  $\mathcal{A}(\cdot)y(\cdot): \Omega \rightarrow Z, \omega \mapsto \mathcal{A}(\omega)y(\omega)$ .

**Corollary 2.7.** Let  $\mathcal{A} \in L^\infty(\Omega, \mathcal{L}(Y, Z))$ . Then the operator  $\mathcal{A}_1: L^\infty(\Omega, Y) \rightarrow L^\infty(\Omega, Z), y(\cdot) \mapsto \mathcal{A}(\cdot)y(\cdot)$  is weakly\*-to-weakly\* continuous, and the affine set  $\{y \in L^\infty(\Omega, Y) : \mathcal{A}(\omega)y(\omega) = b(\omega) \text{ a.s.}\}$  is for all  $b \in L^\infty(\Omega, Z)$  weakly\* closed. Moreover, the operator  $\mathcal{A}_2: Y \rightarrow L^\infty(\Omega, Z), y \mapsto \mathcal{A}(\cdot)y$  is weakly-to-weakly\* continuous.

## 2.3 Risk measures

As the class of problems represented by (1) contains risk measures, we briefly mention some properties that will later be used in the main results. Let  $L^p(\Omega)$  be the space of random variables with  $p \in [1, \infty)$ . Throughout the paper,  $p'$  will denote the Hölder conjugate of  $p$ , i.e.,  $\frac{1}{p} + \frac{1}{p'} = 1$ . Following [4], a risk measure  $\mathcal{R}: L^p(\Omega) \rightarrow \mathbb{R} \cup \{\infty\}$  is called *coherent*, provided it satisfies the following conditions for  $\xi, \xi' \in L^p(\Omega)$ :

- (R1) *Convexity*:  $\mathcal{R}[\lambda\xi + (1 - \lambda)\xi'] \leq \lambda\mathcal{R}[\xi] + (1 - \lambda)\mathcal{R}[\xi'] \quad \forall \lambda \in [0, 1]$ ,
- (R2) *Monotonicity*: If  $\xi \leq \xi'$  a.s., then  $\mathcal{R}[\xi] \leq \mathcal{R}[\xi']$ ,
- (R3) *Translation equivariance*: If  $c \in \mathbb{R}$ , then  $\mathcal{R}[\xi + c] = \mathcal{R}[\xi] + c$ ,
- (R4) *Positive homogeneity*: If  $\lambda > 0$ , then  $\mathcal{R}[\lambda\xi] = \lambda\mathcal{R}[\xi]$ .

Coherent risk measures include, for example,  $\mathcal{R}[\xi] = \mathbb{E}[\xi]$ ,  $\mathcal{R}[\xi] = \sup \xi$ , and average value-at-risk  $\mathcal{R}[\xi] = \text{AVaR}_\alpha[\xi]$  (often called conditional value-at-risk). However, some important risk measures do not satisfy all of these axioms. For the purposes of duality, (R1) and (R2) play a central role, and we



focus on this case. If  $\mathcal{R}$  satisfies (R1) and is proper and lower semicontinuous, then  $\mathcal{R}$  is related to its convex conjugate  $\mathcal{R}^* : L^p(\Omega) \rightarrow \mathbb{R} \cup \{\infty\}$  by

$$\mathcal{R}[\xi] = \sup_{\vartheta \in \text{dom}(\mathcal{R}^*)} \{\mathbb{E}[\xi\vartheta] - \mathcal{R}^*[\vartheta]\}. \quad (4)$$

If  $\mathcal{R}[\xi]$  is also finite, then the maximum is attained in (4) with

$$\partial\mathcal{R}[\xi] = \arg \max_{\vartheta \in \text{dom}(\mathcal{R}^*)} \{\mathbb{E}[\xi\vartheta] - \mathcal{R}^*[\vartheta]\}. \quad (5)$$

It is straightforward to show that if (R2) is satisfied, then  $\vartheta \in \text{dom}(\mathcal{R}^*)$  if and only if  $\vartheta \geq 0$  a.s. Additionally, if  $\mathcal{R} : L^p(\Omega) \rightarrow \mathbb{R}$  satisfies (R1) and (R2) and is everywhere finite, then it is continuous and subdifferentiable on  $L^p(\Omega)$  [41, Proposition 6.5].

While we do not make use of this fact here, it is worth mentioning that the conditions (R3)-(R4) provide a useful characterization of  $\text{dom}(\mathcal{R}^*)$  and the subdifferential. In particular, a risk measure satisfying (R1)-(R3) is equivalent to the restriction  $\vartheta$  to the set of probability density functions (cf. [41, Theorem 6.4]), i.e., the set  $\mathfrak{P} = \{\vartheta \in L^p(\Omega) : \mathbb{E}[\vartheta] = 1, \vartheta \geq 0 \text{ a.s.}\}$ . Finally, a risk measure satisfying (R1)-(R4) implies  $\partial\mathcal{R}[\xi] = \arg \max_{\vartheta \in \mathfrak{P}} \mathbb{E}[\xi\vartheta]$ . For a more thorough introduction to risk measures, we refer to [32, 41].

### 3 Karush–Kuhn–Tucker (KKT) Conditions

We now focus on a particular subclass of (1), namely on convex problems where the set  $X_{2,\text{ad}}(x_1, \omega)$  restricts  $x_2(\omega)$  in the form of an equality and conical constraint. As already mentioned in the introduction, under standard assumptions, classical convex duality theory would provide the existence of Lagrange multipliers in spaces of the form  $(L^\infty(\Omega, X))^*$ . The formulation of pointwise optimality conditions, which we present in this section, is more delicate.

#### 3.1 Problem Formulation

We first equip the problem (1) with more structure. For the remainder of the paper,  $X_1, X_2, W$ , and  $R$  are assumed to be real, reflexive, and separable Banach spaces. We define equality and inequality constraint by the mappings  $e : X_1 \times X_2 \times \Omega \rightarrow W$  and  $i : X_1 \times X_2 \times \Omega \rightarrow R$ , respectively. Given a cone  $K \subset R$ , the partial order  $\leq_K$  is defined by

$$r \leq_K 0 \iff -r \in K,$$

or equivalently,  $r \geq_K 0$  if and only if  $r \in K$ . With

$$x := (x_1, x_2) \in X := X_1 \times L^\infty(\Omega, X_2),$$

we define the problem

$$\begin{aligned} & \underset{x \in X}{\text{minimize}} \quad \{j(x) := J_1(x_1) + \mathcal{R}[J_2(x)]\} \\ & \text{s.t.} \quad \begin{cases} x \in X_0 := \{x \in X : x_1 \in C\}, \\ e(x_1, x_2(\omega); \omega) = 0 \text{ a.s.}, \\ i(x_1, x_2(\omega); \omega) \leq_K 0 \text{ a.s.} \end{cases} \end{aligned} \quad (\text{P})$$

Throughout our work, we often use the notational convention, where if  $x \in X$ ,  $J_2(x)$  is a shorthand for  $J_2(x_1, x_2(\cdot); \cdot)$ . Similarly, if  $x \in X$ , we write  $e(x) = e(x_1, x_2(\cdot); \cdot)$  and  $i(x) = i(x_1, x_2(\cdot); \cdot)$ . Problem (P) is subject to the following assumptions.

**Assumption 3.1.** (i)  $C \subset X_1$  is nonempty, closed, and convex and  $K \subset R$  is a nonempty, closed, and convex cone.

(ii)  $J_1$  is convex, lower semicontinuous, and everywhere finite on  $X_1$ .

(iii)  $J_2(\cdot, \cdot; \omega)$  is continuous and convex with respect to  $X_1 \times X_2$  for almost all  $\omega \in \Omega$  and for every  $(x_1, x_2) \in X_1 \times X_2$ , the functions  $J_2(x_1, x_2; \cdot)$  are measurable on  $\Omega$ . For every  $r > 0$  there exists  $a_r \in L^p(\Omega)$ ,  $p \in [1, \infty)$  such that for almost all  $\omega$  and any  $\|x_1\|_{X_1} + \|x_2\|_{X_2} \leq r$ , we have

$$|J_2(x_1, x_2; \omega)| \leq a_r(\omega). \quad (6)$$

(iv) With respect to  $X_1 \times X_2$ ,  $e(\cdot, \cdot; \omega)$  is continuous and linear and  $i(\cdot, \cdot; \omega)$  is continuous and  $K$ -convex<sup>1</sup> for almost all  $\omega \in \Omega$ .

(v) For every  $(x_1, x_2) \in X_1 \times X_2$ , the functions  $e(x_1, x_2; \cdot)$  and  $i(x_1, x_2; \cdot)$  are measurable on  $\Omega$ . For every  $r > 0$ , there exist constants  $b_{r,e} > 0$  and  $b_{r,i} > 0$  such that for any  $\|x_1\|_{X_1} + \|x_2\|_{X_2} \leq r$ , we have

$$\|e(x_1, x_2; \omega)\|_W \leq b_{r,e}, \quad \|i(x_1, x_2; \omega)\|_R \leq b_{r,i} \quad \text{a.s.} \quad (7)$$

(vi)  $\mathcal{R} : L^p(\Omega) \rightarrow \mathbb{R}$  is convex, monotone, and everywhere finite.

**Remark 3.2.** The growth condition (6) in combination with the Carathéodory property ensures that  $J_2 : X \rightarrow L^p(\Omega)$  is bounded and continuous; see [26, Theorem 3.5]. Boundedness and continuity in the case  $p = \infty$  is possible under additional assumptions; see [3, Theorem 3.17]. The conditions (7) ensure  $e(\chi_1(\cdot), \chi_2(\cdot); \cdot) \in L^\infty(\Omega, W)$  and  $i(\chi_1(\cdot), \chi_2(\cdot); \cdot) \in L^\infty(\Omega, R)$  for all  $(\chi_1, \chi_2) \in L^\infty(\Omega, X_1) \times L^\infty(\Omega, X_2)$ . This structure is used in Theorem 3.9 in connection with the nonanticipativity constraint.

**Remark 3.3.** Suppose  $h : X_1 \times X_2 \times \Omega \rightarrow \mathbb{R}$  is a normal convex integrand and  $h(x_1, x_2(\omega); \omega)$  is an integrable function of  $\omega$  for every  $x \in X$ . Then  $x \mapsto I_h(x)$  is a weakly\* lower semicontinuous mapping on  $X$ . This can be seen by pairing  $X$  with  $X' := X_1^* \times L^1(\Omega, X_2^*)$  and arguing that  $I_h$  and  $I_{h^*}$  are well-defined on  $X$  and  $X'$ , respectively. These functionals are conjugate to each other as in Lemma 2.1 with

$$I_{h^*}(x') = \sup_{x \in X} \{ \langle x', x \rangle_{X', X} - I_h(x) \} \quad \text{and} \quad I_h(x) = \sup_{x' \in X'} \{ \langle x', x \rangle_{X', X} - I_{h^*}(x') \}.$$

This gives weak\* lower semicontinuity using the arguments in Remark 2.2. Here, the weak\* topology coincides with the weak topology on (the reflexive space)  $X_1$ , so one even has weak-weak\* lower semicontinuity of  $I_h$ , i.e., weak with respect to  $X_1$  and weak\* with respect to  $L^\infty(\Omega, X_2)$ . To simplify the discussion, we will not make the distinction between weak and weak\* with respect to  $x_1$ .

The following result is an adaptation of [26, Proposition 3.8].

**Proposition 3.4.** Suppose Assumption 3.1(iii) holds and  $\mathcal{R} : L^p(\Omega) \rightarrow \mathbb{R} \cup \{\infty\}$  is proper, convex, monotone, and lower semicontinuous. If  $\mathcal{R}$  is finite at  $J_2(x)$ , then  $\mathcal{R} \circ J_2$  is weakly\* lower semicontinuous at  $x$ .

<sup>1</sup> $K$ -convexity of  $i(\cdot, \cdot; \omega)$  means

$$i(\lambda x_1 + (1 - \lambda)\hat{x}_1, \lambda x_2 + (1 - \lambda)\hat{x}_2; \omega) \leq_K \lambda i(x_1, x_2; \omega) + (1 - \lambda)i(\hat{x}_1, \hat{x}_2; \omega) \quad \forall (x_1, x_2), (\hat{x}_1, \hat{x}_2) \in X_1 \times X_2.$$

*Proof.* Let  $f(x_1, x_2; \omega) := J_2(x_1, x_2; \omega)\vartheta(\omega)$ , where  $\vartheta \in \text{dom}(\mathcal{R}^*) \subset L^{p'}(\Omega)$ . This function is clearly a normal convex integrand since  $\vartheta \geq 0$  a.s. by monotonicity of  $\mathcal{R}$ . Moreover (6) in combination with  $\vartheta \in L^{p'}(\Omega)$  imply that  $f$  is integrable on  $X$ . Weak\* lower semicontinuity of  $I_f$  on  $X$  follows by Remark 3.3.

Now, suppose  $\{x^n\} \subset X$  is a sequence such that  $x^n \rightharpoonup^* x$ . Then

$$\mathcal{R}[J_2(x^n)] = \sup_{\vartheta \in \text{dom}(\mathcal{R}^*)} \{\mathbb{E}[J_2(x^n)\vartheta] - \mathcal{R}^*[\vartheta]\} \geq \mathbb{E}[J_2(x^n)\theta] - \mathcal{R}^*[\theta] \quad (8)$$

for any  $\theta \in \text{dom}(\mathcal{R}^*)$ . Since  $\mathcal{R}$  is finite at  $J_2(x)$ , it is subdifferentiable there, so (8) holds for  $\theta \in \partial\mathcal{R}[J_2(x)]$ . With (5) and weak\* lower semicontinuity of  $I_f$  on  $X$ , weak\* lower semicontinuity of  $\mathcal{R} \circ J_2$  at  $x$  follows.  $\square$

As a result of Proposition 3.4, Assumption 3.1(iii), and Assumption 3.1(vi) imply weak\* lower semicontinuity of  $\mathcal{R} \circ J_2$  on  $X$ .

## 3.2 Definition of Lagrangians

To obtain pointwise optimality conditions, it will be advantageous to use the decomposition afforded by Theorem 2.4. This structure will be used to distinguish between the extended (standard) Lagrangian associated with the natural pairing and the Lagrangian associated with the weak\* pairing. We define the sets  $\Lambda := L^1(\Omega, W^*) \times L^1(\Omega, R^*)$  and  $\Lambda^\circ := \{\lambda^\circ = (\lambda_e^\circ, \lambda_i^\circ) \in \mathcal{S}_e \times \mathcal{S}_i\}$ , where  $\mathcal{S}_e$  and  $\mathcal{S}_i$  denote the sets of singular functionals (as in Definition 2.3) defined on  $L^\infty(\Omega, W)$  and  $L^\infty(\Omega, R)$ , respectively. We define the dual cones

$$K^\oplus := \{r^* \in R^* : \langle r^*, r \rangle_{R^*, R} \geq 0 \quad \forall r \in K\}.$$

$$\mathcal{K}^\oplus := \{\lambda_i^\circ \in \mathcal{S}_i : \langle \lambda_i^\circ, y \rangle_{(L^\infty(\Omega, R))^*, L^\infty(\Omega, R)} \geq 0 \quad \forall y \in L^\infty(\Omega, R) \text{ satisfying } y(\omega) \in K \text{ a.s.}\}.$$

Now, we define the admissible sets  $\Lambda_0 = \{\lambda = (\lambda_e, \lambda_i) \in \Lambda : \lambda_i(\omega) \in K^\oplus \text{ a.s.}\}$  and  $\Lambda_0^\circ = \{\lambda^\circ = (\lambda_e^\circ, \lambda_i^\circ) \in \Lambda^\circ : \lambda_i^\circ \in \mathcal{K}^\oplus\}$  corresponding to the associated dual problem.

Finally, we define the *extended Lagrangian* over  $X_0 \times \Lambda_0 \times \Lambda_0^\circ$  by

$$\bar{L}(x, \lambda, \lambda^\circ) = L(x, \lambda) + L^\circ(x, \lambda^\circ), \quad (9)$$

where

$$\begin{aligned} L(x, \lambda) &:= j(x) + \langle \lambda_e, e(x) \rangle_{L^1(\Omega, W^*), L^\infty(\Omega, W)} + \langle \lambda_i, i(x) \rangle_{L^1(\Omega, R^*), L^\infty(\Omega, R)} \\ &= j(x) + \mathbb{E}[\langle \lambda_e, e(x) \rangle_{W^*, W} + \langle \lambda_i, i(x) \rangle_{R^*, R}] \end{aligned}$$

is called the *Lagrangian* and

$$L^\circ(x, \lambda^\circ) := \langle \lambda_e^\circ, e(x) \rangle_{(L^\infty(\Omega, W))^*, L^\infty(\Omega, W)} + \langle \lambda_i^\circ, i(x) \rangle_{(L^\infty(\Omega, R))^*, L^\infty(\Omega, R)}.$$

By convention, we set  $\bar{L}(x, \lambda, \lambda^\circ) = -\infty$  if  $x \in X_0$  but  $(\lambda, \lambda^\circ) \notin \Lambda_0 \times \Lambda_0^\circ$ , and  $\bar{L}(x, \lambda, \lambda^\circ) = \infty$  if  $x \notin X_0$ .

By Theorem 2.4,  $\lambda_e$  and  $\lambda_i$  can be identified with elements  $\lambda_e^a$  and  $\lambda_i^a$  of the spaces  $(L^\infty(\Omega, W))^*$  and  $(L^\infty(\Omega, R))^*$ , respectively. Hence (9) simply corresponds to the standard Lagrangian when

$$U := L^\infty(\Omega, W) \times L^\infty(\Omega, R)$$

is paired with its dual  $U^*$ .

We will now present the conditions under which saddle points to the extended Lagrangian  $\bar{L}$  exist. To that end, let  $u = (u_e, u_i) \in U$  be a point in a neighborhood of zero in  $U$ . We define the perturbed admissible set

$$F_{\text{ad},u} := \{x \in X : x_1 \in C, e(x_1, x_2(\omega); \omega) = u_e(\omega) \text{ a.s.}, i(x_1, x_2(\omega); \omega) \leq_K u_i(\omega) \text{ a.s.}\}$$

and the value function

$$v(u) := \inf_{x \in X} j(x) + \delta_{F_{\text{ad},u}}(x).$$

For the existence of saddle points, we impose the following standard assumption.

**Assumption 3.5.** (i)  $F_{\text{ad},u}$  is bounded for all  $u$  in a neighborhood of zero or  $j$  is radially unbounded, meaning  $\|x\|_X \rightarrow \infty$  implies  $j(x) \rightarrow \infty$ .

(ii) The problem is strictly feasible in the sense that

$$0 \in \text{int dom } v. \quad (10)$$

The condition (10) is satisfied if  $v$  is finite in a neighborhood of zero and corresponds to an “almost sure”-type Slater condition.

To obtain “regular” Lagrange multipliers, meaning multipliers with respect to the Lagrangian  $L$ , we make an additional assumption.

**Assumption 3.6** (relatively complete recourse). The induced feasible set

$$\tilde{C} := \{x_1 \in X_1 : \exists x_2 \in L^\infty(\Omega, X_2) \text{ satisfying } e(x_1, x_2(\omega); \omega) = 0 \text{ a.s.}, i(x_1, x_2(\omega); \omega) \leq_K 0 \text{ a.s.}\}$$

contains the first-stage feasible set, i.e.  $C \subset \tilde{C}$ .

Assumption 3.6 is quite strong, essentially requiring the second-stage variable to be feasible for any feasible first-stage variable. In Section 3.4, we will see that for problems satisfying this condition, Lagrange multipliers happen to be more regular. In that case, integrable multipliers in  $\Lambda$  can be found and the singular terms from  $\Lambda^\circ$  vanish.

### 3.3 KKT conditions

In this section, we present strong KKT conditions without the additional Assumption 3.6. This leads to Lagrange multipliers belonging to  $\Lambda \times \Lambda^\circ$ , i.e., multipliers have singular parts. Theorem 3.9 uses arguments made in [38, Section 5], where the finite-dimensional case was considered. Many arguments from [19] can be carried over; we repeat some arguments here and in Appendix B for completeness. First, we obtain the existence of saddle points.

**Lemma 3.7.** Let Assumption 3.1 and Assumption 3.5 be satisfied. Then there exists a saddle point  $(\bar{x}, \bar{\lambda}, \bar{\lambda}^\circ) \in X_0 \times \Lambda_0 \times \Lambda_0^\circ$  to the Lagrangian  $\bar{L}$ , i.e, the point  $(\bar{x}, \bar{\lambda}, \bar{\lambda}^\circ)$  satisfies

$$\bar{L}(\bar{x}, \lambda, \lambda^\circ) \leq \bar{L}(\bar{x}, \bar{\lambda}, \bar{\lambda}^\circ) \leq \bar{L}(x, \bar{\lambda}, \bar{\lambda}^\circ) \quad \forall (x, \lambda, \lambda^\circ) \in X \times \Lambda \times \Lambda^\circ. \quad (11)$$

*Proof.* This is proven in Theorem B.2 and B.3. □

It will be convenient to define the function

$$\ell(x_1, \lambda^\circ) = \inf_{x_2 \in L^\infty(\Omega, X_2)} L^\circ(x, \lambda^\circ). \quad (12)$$

Let us start with a technical lemma, which adapts [39, Theorem 3] to include risk functionals.

**Lemma 3.8.** Let Assumption 3.1 be satisfied. Then for every  $(\lambda, \lambda^\circ) \in \Lambda_0 \times \Lambda_0^\circ$ , we have

$$\inf_{x \in X_0} \{L(x, \lambda) + L^\circ(x, \lambda^\circ)\} = \inf_{x \in X_0} \{L(x, \lambda) + \ell(x_1, \lambda^\circ)\}.$$

*Proof.* To begin, notice that for all  $x_1 \in C_1$ , we trivially have

$$\inf_{x_2 \in L^\infty(\Omega, X_2)} \{L(x, \lambda) + L^\circ(x, \lambda^\circ)\} \geq \inf_{x_2 \in L^\infty(\Omega, X_2)} L(x, \lambda) + \inf_{x_2 \in L^\infty(\Omega, X_2)} L^\circ(x, \lambda^\circ).$$

Hence it is enough to show the inequality “ $\leq$ ”, or equivalently, that for arbitrary  $\varepsilon > 0$  and arbitrarily fixed  $x' := (x_1, y')$ ,  $x'' := (x_1, y'') \in X_0$ , there exists  $x = (x_1, y) \in X_0$  such that the inequality

$$L(x, \lambda) + L^\circ(x, \lambda^\circ) \leq L(x'', \lambda) + L^\circ(x', \lambda^\circ) + \varepsilon \quad (13)$$

is true. Then (13) in combination with the definition (12) will prove the assertion.

Let  $\{F_{e,n}\}$  and  $\{F_{i,n}\}$  be the sets on which  $\lambda_e^\circ$  and  $\lambda_i^\circ$  are concentrated according to Definition 2.3. We define  $F_n = F_{e,n} \cup F_{i,n}$  and

$$y^n(\omega) = \begin{cases} y'(\omega), & \omega \in F_n \\ y''(\omega), & \omega \notin F_n \end{cases}.$$

On  $F_n$ , we have  $e(x_1, y^n(\omega); \omega) = e(x_1, y'(\omega); \omega)$  and  $i(x_1, y^n(\omega); \omega) = i(x_1, y'(\omega); \omega)$ . Therefore, thanks to Definition 2.3, for  $x^n = (x_1, y^n)$  we have

$$\langle \lambda_e^\circ, e(x^n) - e(x') \rangle_{(L^\infty(\Omega, W))^*, L^\infty(\Omega, W)} = 0, \quad \langle \lambda_i^\circ, i(x^n) - i(x') \rangle_{(L^\infty(\Omega, R))^*, L^\infty(\Omega, R)} = 0,$$

which implies

$$L^\circ(x^n, \lambda^\circ) = L^\circ(x', \lambda^\circ) \quad \forall n. \quad (14)$$

Let  $\mathbb{1}_A$  denote the indicator function on  $A$ , defined by  $\mathbb{1}_A(x) = 1$  if  $x \in A$  and  $\mathbb{1}_A(x) = 0$  otherwise. Then  $J_2(x_1, y^n(\cdot); \cdot) \rightarrow J_2(x_1, y''(\cdot); \cdot)$  in  $L^p(\Omega)$  as  $n \rightarrow \infty$ . Indeed,

$$\begin{aligned} & \|J_2(x_1, y^n(\cdot); \cdot) - J_2(x_1, y''(\cdot); \cdot)\|_{L^p(\Omega)}^p \\ &= \int_{\Omega} |J_2(x_1, y''(\omega); \omega) \mathbb{1}_{\Omega \setminus F_n}(\omega) + J_2(x_1, y'(\omega); \omega) \mathbb{1}_{F_n}(\omega) - J_2(x_1, y''(\omega); \omega)|^p \, d\mathbb{P}(\omega) \\ &= \int_{\Omega \setminus F_n} |J_2(x_1, y''(\omega); \omega) - J_2(x_1, y''(\omega); \omega)|^p \, d\mathbb{P}(\omega) \\ &\quad + \int_{F_n} |J_2(x_1, y'(\omega); \omega) - J_2(x_1, y''(\omega); \omega)|^p \, d\mathbb{P}(\omega). \end{aligned}$$

Since the integrand in the last integral above is bounded and  $\mathbb{P}(F_n) \rightarrow 0$ , the last integral vanishes. We now show that  $x^n$  converges strongly to  $x''$  in  $X$ . Notice that

$$\begin{aligned} \|x^n - x''\|_X &= \|y^n - y''\|_{L^\infty(\Omega, X_2)} = \operatorname{ess\,sup}_{\omega \in \Omega} \|y^n(\omega) - y''(\omega)\|_{X_2} \\ &= \operatorname{ess\,sup}_{\omega \in F_n} \|y^n(\omega) - y''(\omega)\|_{X_2} \\ &= \inf\{a \in \mathbb{R} : \mathbb{P}\{\omega \in F_n : \|y^n(\omega) - y''(\omega)\|_{X_2} > a\} = 0\}. \end{aligned}$$

In combination with  $\mathbb{P}(F_n) \rightarrow 0$  as  $n \rightarrow \infty$ , the event  $\{\omega \in F_n : \|y^n(\omega) - y''(\omega)\|_{X_2} > a\}$  has vanishing small probability for all  $a \geq 0$  as  $n \rightarrow \infty$ , i.e., in particular for the choice  $a = 0$ . Recall that  $J_2: X \rightarrow L^p(\Omega)$  is continuous by Remark 3.2. Now, continuity of  $\mathcal{R}: L^p(\Omega) \rightarrow \mathbb{R}$  gives

$$\lim_{n \rightarrow \infty} \mathcal{R}[J_2(x^n)] = \mathcal{R}[J_2(x'')].$$

Analogous arguments yield

$$\lim_{n \rightarrow \infty} \mathbb{E}[\langle \lambda_e, e(x^n) \rangle_{W^*, W}] = \mathbb{E}[\langle \lambda_e, e(x'') \rangle_{W^*, W}], \quad \lim_{n \rightarrow \infty} \mathbb{E}[\langle \lambda_i, i(x^n) \rangle_{R^*, R}] = \mathbb{E}[\langle \lambda_i, i(x'') \rangle_{R^*, R}].$$

As a result, we obtain  $\lim_{n \rightarrow \infty} L(x^n, \lambda) = L(x'', \lambda)$ . In combination with (14), for any  $\varepsilon > 0$ , there exists an  $n_0$  such that for all  $n \geq n_0$  we have

$$L(x^n, \lambda) + L^\circ(x^n, \lambda^\circ) \leq L(x'', \lambda) + L^\circ(x', \lambda^\circ) + \varepsilon.$$

After setting  $x := x^n$  (for an arbitrary  $n \geq n_0$ ), we obtain (13).  $\square$

**Theorem 3.9.** Let Assumption 3.1 be satisfied. Then  $(\bar{x}, \bar{\lambda}, \bar{\lambda}^\circ) \in X_0 \times \Lambda_0 \times \Lambda_0^\circ$  satisfies (11) if and only if there exist  $\bar{\rho} \in L^1(\Omega, X_1^*)$  and  $\bar{\vartheta} \in \partial \mathcal{R}[J_2(\bar{x})]$  fulfilling the following conditions:

(i) The function

$$x_1 \mapsto J_1(x_1) + \langle \mathbb{E}[\bar{\rho}], x_1 \rangle_{X_1^*, X_1} + \ell(x_1, \bar{\lambda}^\circ)$$

attains its minimum over  $C$  at  $\bar{x}_1$ .

(ii) The function

$$(x_1, x_2) \mapsto J_2(x_1, x_2; \omega) \bar{\vartheta}(\omega) + \langle \bar{\lambda}_e(\omega), e(x_1, x_2; \omega) \rangle_{W^*, W} \\ + \langle \bar{\lambda}_i(\omega), i(x_1, x_2; \omega) \rangle_{R^*, R} - \langle \bar{\rho}(\omega), x_1 \rangle_{X_1^*, X_1}$$

attains its minimum in  $X_1 \times X_2$  at  $(\bar{x}_1, \bar{x}_2(\omega))$  for almost every  $\omega \in \Omega$ .

(iii) It holds that  $\bar{x}_1 \in C$  and the following conditions hold almost surely:

$$e(\bar{x}_1, \bar{x}_2(\omega); \omega) = 0, \\ \bar{\lambda}_i(\omega) \in K^\oplus, \quad i(\bar{x}_1, \bar{x}_2(\omega); \omega) \leq_K 0, \quad \langle \bar{\lambda}_i(\omega), i(\bar{x}_1, \bar{x}_2(\omega); \omega) \rangle_{R^*, R} = 0.$$

(iv) It holds that  $\ell(\bar{x}_1, \bar{\lambda}^\circ) = 0$  and

$$\bar{\lambda}_i^\circ \in \mathcal{K}^\oplus, \quad \langle \bar{\lambda}_i^\circ, i(\bar{x}) \rangle_{(L^\infty(\Omega, R))^*, L^\infty(\Omega, R)} = 0.$$

The origin of the multiplier  $\bar{\rho}$  in Theorem 3.9 will become apparent in the proof. There, in view of applying the Moreau–Rockafellar theorem, the variable  $x_1 \in X_1$  first needs to be identified with its corresponding element in the Bochner space  $L^\infty(\Omega, X_1)$ . This identification is equivalent to imposing  $x_1(\omega) \equiv x_1$ , meaning  $x_1(\omega) = x_1$  a.s. Sometimes, this condition is imposed explicitly in the problem formulation and is known as a “nonanticipativity” constraint; see [41, Section 2.4].

**Remark 3.10.** Recall the convention  $\bar{L}(x, \lambda, \lambda^\circ) = -\infty$  if  $x \in X_0$  and  $(\lambda, \lambda^\circ) \notin \Lambda_0 \times \Lambda_0^\circ$ . Given  $\lambda^\circ \in \Lambda^\circ$ , we have the implication

$$e(\bar{x}_1, \bar{x}_2(\omega); \omega) = 0 \text{ a.s.}, \quad i(\bar{x}_1, \bar{x}_2(\omega); \omega) \leq_K 0 \text{ a.s.} \quad \Rightarrow \quad L^\circ(\bar{x}, \lambda^\circ) \leq 0. \quad (15)$$

Hence conditions (iii) and (iv) imply

$$0 = \ell(\bar{x}_1, \bar{\lambda}^\circ) \leq L^\circ(\bar{x}, \bar{\lambda}^\circ) \leq 0. \quad (16)$$

*Proof of Theorem 3.9.* We show that if conditions (i)–(iv) are satisfied, then a saddle point exists. Notice that  $\bar{\vartheta} \in \partial\mathcal{R}[J_2(\bar{x})]$ , (4), and (5) imply with  $\xi = J_2(\bar{x})$

$$\mathbb{E}[J_2(x)\bar{\vartheta}] - \mathcal{R}^*[\bar{\vartheta}] = \mathcal{R}[J_2(\bar{x})]. \quad (17)$$

Additionally, (4) implies with  $\xi = J_2(x)$

$$\mathbb{E}[J_2(x)\bar{\vartheta}] - \mathcal{R}^*[\bar{\vartheta}] \leq \sup_{\vartheta \in \text{dom}(\mathcal{R}^*)} \{\mathbb{E}[J_2(x)\vartheta] - \mathcal{R}^*[\vartheta]\} = \mathcal{R}[J_2(x)]. \quad (18)$$

Now,

$$\begin{aligned} \bar{L}(\bar{x}, \bar{\lambda}, \bar{\lambda}^\circ) &= L(\bar{x}, \bar{\lambda}) + L^\circ(\bar{x}, \bar{\lambda}^\circ) \\ &\stackrel{(16)}{=} J_1(\bar{x}_1) + \mathcal{R}[J_2(\bar{x})] + \mathbb{E}[\langle \bar{\lambda}_e, e(\bar{x}) \rangle_{W^*, W} + \langle \bar{\lambda}_i, i(\bar{x}) \rangle_{R^*, R}] + \ell(\bar{x}_1, \bar{\lambda}^\circ) \\ &\stackrel{(17)}{=} J_1(\bar{x}_1) + \ell(\bar{x}_1, \bar{\lambda}^\circ) + \mathbb{E}[J_2(\bar{x})\bar{\vartheta}] - \mathcal{R}^*[\bar{\vartheta}] + \mathbb{E}[\langle \bar{\lambda}_e, e(\bar{x}) \rangle_{W^*, W} + \langle \bar{\lambda}_i, i(\bar{x}) \rangle_{R^*, R}] \\ &= J_1(\bar{x}_1) + \langle \mathbb{E}[\bar{\rho}], \bar{x}_1 \rangle_{X_1^*, X_1} + \ell(\bar{x}_1, \bar{\lambda}^\circ) + \mathbb{E}[J_2(\bar{x})\bar{\vartheta}] - \mathcal{R}^*[\bar{\vartheta}] \\ &\quad + \mathbb{E}[\langle \bar{\lambda}_e, e(\bar{x}) \rangle_{W^*, W} + \langle \bar{\lambda}_i, i(\bar{x}) \rangle_{R^*, R} - \langle \bar{\rho}, \bar{x}_1 \rangle_{X_1^*, X_1}] \\ &\stackrel{(i), (ii)}{\leq} J_1(x_1) + \langle \mathbb{E}[\bar{\rho}], x_1 \rangle_{X_1^*, X_1} + \ell(x_1, \bar{\lambda}^\circ) + \mathbb{E}[J_2(x)\bar{\vartheta}] - \mathcal{R}^*[\bar{\vartheta}] \\ &\quad + \mathbb{E}[\langle \bar{\lambda}_e, e(x) \rangle_{W^*, W} + \langle \bar{\lambda}_i, i(x) \rangle_{R^*, R} - \langle \bar{\rho}, x_1 \rangle_{X_1^*, X_1}] \\ &\stackrel{(18)}{\leq} \bar{L}(x, \bar{\lambda}, \bar{\lambda}^\circ) \quad \forall x \in X. \end{aligned}$$

Note that we used monotonicity of the expectation operator in the second to last inequality above.

To prove that  $\bar{L}(\bar{x}, \lambda, \lambda^\circ) \leq \bar{L}(\bar{x}, \bar{\lambda}, \bar{\lambda}^\circ)$  for all  $(\lambda, \lambda^\circ) \in \Lambda \times \Lambda^\circ$ , we first prove  $L(\bar{x}, \lambda) \leq L(\bar{x}, \bar{\lambda})$  for all  $\lambda \in \Lambda$ . Notice this follows if

$$\mathbb{E}[\langle \lambda_e, e(\bar{x}) \rangle_{W^*, W} + \langle \lambda_i, i(\bar{x}) \rangle_{R^*, R}] \leq \mathbb{E}[\langle \bar{\lambda}_e, e(\bar{x}) \rangle_{W^*, W} + \langle \bar{\lambda}_i, i(\bar{x}) \rangle_{R^*, R}],$$

for all  $\lambda \in \Lambda$ . This is fulfilled thanks to condition (iii) and the fact that  $\langle \lambda_i(\omega), i(\bar{x}_1, \bar{x}_2(\omega); \omega) \rangle_{R^*, R} \leq 0$  a.s. for all  $\lambda \in \Lambda$ . Now, it is enough to argue that  $L^\circ(\bar{x}, \lambda^\circ) \leq L^\circ(\bar{x}, \bar{\lambda}^\circ)$  for all  $\lambda^\circ \in \Lambda^\circ$ . This is also clear, since given  $\lambda^\circ \in \Lambda^\circ$  and conditions (iii) and (iv), (16) holds, hence  $0 = L^\circ(\bar{x}, \bar{\lambda}^\circ) = \sup_{\lambda^\circ \in \Lambda^\circ} L^\circ(\bar{x}, \lambda^\circ) \geq L^\circ(\bar{x}, \lambda^\circ)$ .

For the second part of the proof, we show that  $(\bar{x}, \bar{\lambda}, \bar{\lambda}^\circ)$  being a saddle point implies conditions (i)–(iv). Trivially,  $(\bar{x}, \bar{\lambda}, \bar{\lambda}^\circ) \in X_0 \times \Lambda_0 \times \Lambda_0^\circ$  implies  $\bar{x}_1 \in C$ ,  $\bar{\lambda}_i(\omega) \in K^\oplus$  a.s., and  $\bar{\lambda}_i^\circ \in \mathcal{K}^\oplus$ . It is straightforward to argue that  $e(\bar{x}_1, \bar{x}_2(\omega); \omega) = 0$  a.s. and  $i(\bar{x}_1, \bar{x}_2(\omega); \omega) \leq_K 0$  a.s., since otherwise one would have  $\sup_{\lambda \in \Lambda} L(\bar{x}, \lambda) = \infty$ . Now,

$$0 \geq L^\circ(\bar{x}, \bar{\lambda}^\circ) = \sup_{\lambda^\circ \in \Lambda_0^\circ} \{ \langle \lambda_e^\circ, e(\bar{x}) \rangle_{(L^\infty(\Omega, W))^*, L^\infty(\Omega, W)} + \langle \lambda_i^\circ, i(\bar{x}) \rangle_{(L^\infty(\Omega, R))^*, L^\infty(\Omega, R)} \}.$$

Since  $0 \in \mathcal{K}^\oplus$ , we obtain

$$L^\circ(\bar{x}, \bar{\lambda}^\circ) = 0$$

and thus  $\ell(\bar{x}_1, \bar{\lambda}^\circ) = 0$  and  $\langle \lambda_i^\circ, i(\bar{x}) \rangle_{(L^\infty(\Omega, R))^*, L^\infty(\Omega, R)} = 0$ . Analogous arguments can be made to conclude that  $\langle \bar{\lambda}_i(\omega), i(\bar{x}_1, \bar{x}_2(\omega); \omega) \rangle_{R^*, R} = 0$  a.s. We have shown that conditions (iii)–(iv) hold.

Now, we show that conditions (i)–(ii) must hold. Since  $(\bar{x}, \bar{\lambda}, \bar{\lambda}^\circ)$  is a saddle point, we have in combination with Lemma 3.8 that

$$\bar{L}(\bar{x}, \bar{\lambda}, \bar{\lambda}^\circ) = \inf_{x \in X_0} \bar{L}(x, \bar{\lambda}, \bar{\lambda}^\circ) = \inf_{x \in X_0} \{L(x, \bar{\lambda}) + \ell(x_1, \bar{\lambda}^\circ)\}. \quad (19)$$

We first work on rewriting the expression  $L(x, \bar{\lambda}) + \ell(x_1, \bar{\lambda}^\circ)$ . We define the convex function

$$H_1(x) = \begin{cases} J_1(x_1) + \ell(x_1, \bar{\lambda}^\circ), & x \in X_0, \\ \infty, & \text{else.} \end{cases} \quad (20)$$

Additionally, we define the integrand

$$h_2(x_1, x_2; \omega) = J_2(x_1, x_2; \omega) \bar{\vartheta}(\omega) + \langle \bar{\lambda}_e(\omega), e(x_1, x_2; \omega) \rangle_{W^*, W} \\ + \langle \bar{\lambda}_i(\omega), i(x_1, x_2; \omega) \rangle_{R^*, R} - \mathcal{R}^*[\bar{\vartheta}].$$

Assumption 3.1 ensures that  $h_2$  is a Carathéodory, i.e., continuous in  $(x_1, x_2)$  for every  $\omega$  and measurable in  $\omega$  for every  $(x_1, x_2)$ . Additionally,  $h_2$  is convex with respect to  $(x_1, x_2)$ ; recall that the monotonicity of  $\mathcal{R}$  implies  $\bar{\vartheta} \geq 0$  a.s. Additionally,  $\bar{\lambda} \in \Lambda_0$  implies  $\bar{\lambda}_i(\omega) \in K^\oplus$  a.s., giving convexity of the mapping  $(x_1, x_2) \mapsto \langle \bar{\lambda}_i(\omega), i(x_1, x_2; \omega) \rangle_{R^*, R}$ . As a result,  $h_2$  is a normal convex integrand. Let  $\mathfrak{X} := L^\infty(\Omega, X_1) \times L^\infty(\Omega, X_2)$ . The growth conditions from Assumption 3.1 ensure that for any  $r > 0$ , there exist  $a_r \in L^p(\Omega)$  and positive constants  $b_{r,e}$  and  $b_{r,i}$  such that for all  $\chi = (\chi_1, \chi_2) \in \mathfrak{X}$  satisfying  $\|\chi\|_{\mathfrak{X}} \leq r$ , we have

$$|h_2(\chi_1(\omega), \chi_2(\omega); \omega)| \leq |J_2(\chi_1(\omega), \chi_2(\omega); \omega) \bar{\vartheta}(\omega)| + |\langle \bar{\lambda}_e(\omega), e(\chi_1(\omega), \chi_2(\omega); \omega) \rangle_{W^*, W}| \\ + |\langle \bar{\lambda}_i(\omega), i(\chi_1(\omega), \chi_2(\omega); \omega) \rangle_{R^*, R}| + |\mathcal{R}^*[\bar{\vartheta}]| \\ \leq |a_r(\omega)| |\bar{\vartheta}(\omega)| + b_{r,e} \|\bar{\lambda}_e(\omega)\|_{W^*} + b_{r,i} \|\bar{\lambda}_i(\omega)\|_{R^*} + |\mathcal{R}^*[\bar{\vartheta}]| =: k_r(\omega).$$

It is straightforward to show that  $k_r \in L^1(\Omega)$ . Since  $r$  is arbitrary,  $h_2(\chi_1(\omega), \chi_2(\omega); \omega)$  is integrable for all  $\chi = (\chi_1, \chi_2) \in \mathfrak{X}$ . In particular, the functional

$$H_2(\chi) = \int_{\Omega} J_2(\chi_1(\omega), \chi_2(\omega); \omega) \bar{\vartheta}(\omega) + \langle \bar{\lambda}_e(\omega), e(\chi_1(\omega), \chi_2(\omega); \omega) \rangle_{W^*, W} \, d\mathbb{P}(\omega) \\ + \int_{\Omega} \langle \bar{\lambda}_i(\omega), i(\chi_1(\omega), \chi_2(\omega); \omega) \rangle_{R^*, R} \, d\mathbb{P}(\omega) - \mathcal{R}^*[\bar{\vartheta}]$$

is well-defined and finite on  $\mathfrak{X}$  by Lemma 2.1. We now prove continuity of  $H_2: \mathfrak{X} \rightarrow \mathbb{R}$  by contradiction. Suppose  $\{\chi^n\}$  is a sequence converging to  $\chi$  in  $\mathfrak{X}$  where  $H_2(\chi^n)$  does not converge to  $H_2(\chi)$ . Then there exists  $\varepsilon > 0$  and a subsequence  $\{\chi^{n_k}\}$  such that

$$|H_2(\chi^{n_k}) - H_2(\chi)| > \varepsilon. \quad (21)$$

On a further subsequence (where we use the same labeling) we have that  $\chi^{n_k}$  converges pointwise in  $\omega$  to  $\chi$  a.s. By continuity of  $(x_1, x_2) \mapsto h_2(x_1, x_2; \omega)$ , we have  $h_2(\chi_1^{n_k}(\omega), \chi_2^{n_k}(\omega); \omega) \rightarrow h_2(\chi_1(\omega), \chi_2(\omega); \omega)$  a.s. Choosing  $r$  large enough, it is possible to construct a  $k_r \in L^1(\Omega)$  as before such that

$$|h_2(\chi_1^{n_k}(\omega), \chi_2^{n_k}(\omega); \omega)| \leq k_r(\omega) \quad \text{a.s.}$$

Therefore by Lebesgue's Dominated Convergence theorem, it follows that  $H_2(\chi^{n_k}) \rightarrow H_2(\chi)$ , a contradiction to (21), proving our claim about continuity of  $H_2$ . Let  $\iota: X \rightarrow \mathfrak{X}$  be the continuous injection, which maps elements of  $X_1$  to the corresponding constant in  $L^\infty(\Omega, X_1)$  and maps each element of  $L^\infty(\Omega, X_2)$  to itself. In particular,  $\iota(\bar{x}) = (\bar{\chi}_1(\cdot), \bar{\chi}_2(\cdot))$  with  $(\bar{\chi}_1(\omega), \bar{\chi}_2(\omega)) = (\bar{x}_1, \bar{x}_2(\omega))$  a.s., implying

$$H_2(\iota(\bar{x})) = \int_{\Omega} J_2(\bar{x}_1, \bar{x}_2(\omega); \omega) \bar{\vartheta}(\omega) + \langle \bar{\lambda}_e(\omega), e(\bar{x}_1, \bar{x}_2(\omega); \omega) \rangle_{W^*, W} \, d\mathbb{P}(\omega) \\ + \int_{\Omega} \langle \bar{\lambda}_i(\omega), i(\bar{x}_1, \bar{x}_2(\omega); \omega) \rangle_{R^*, R} \, d\mathbb{P}(\omega) - \mathcal{R}^*[\bar{\vartheta}] \quad (22) \\ = \mathcal{R}[J_2(\bar{x})] + \mathbb{E}[\langle \lambda_e, e(\bar{x}) \rangle_{W^*, W} + \langle \lambda_i, i(\bar{x}) \rangle_{R^*, R}],$$



where the last equality follows since  $\bar{\vartheta} \in \partial\mathcal{R}[J_2(\bar{x})]$ .

Combining (19), (20) and (22), we have

$$\inf_{x \in X_0} \bar{L}(x, \bar{\lambda}, \bar{\lambda}^\circ) = \bar{L}(\bar{x}, \bar{\lambda}, \bar{\lambda}^\circ) = H_1(\bar{x}) + H_2(\iota(\bar{x})).$$

Since  $H_2$  is continuous on  $\mathfrak{X}$  and  $\bar{x} \in \text{dom}(H_1)$  (note  $x \in X_0$  implies  $x_1 \in C$ ), it is possible to invoke the Moreau–Rockafellar theorem (cf. [10, Theorem 2.168]) to obtain

$$0 \in \partial H_1(\bar{x}) + \iota^* \partial H_2(\iota(\bar{x})).$$

In particular, there exists  $\bar{q} \in \mathfrak{X}^* = (L^\infty(\Omega, X_1) \times L^\infty(\Omega, X_2))^*$  such that

$$-\iota^* \bar{q} \in \partial H_1(\bar{x}) \quad \text{and} \quad \bar{q} \in \partial H_2(\iota(\bar{x})).$$

From Lemma 2.5, it follows that  $\partial H_2(\iota(\bar{x}))$  consists of continuous linear functionals on  $L^\infty(\Omega, X_1) \times L^\infty(\Omega, X_2)$ , which can be identified with pairs  $(\bar{q}_1, \bar{q}_2)$  from a weakly compact subset of  $L^1(\Omega, X_1^*) \times L^1(\Omega, X_2^*)$  such that

$$\bar{q}(\omega) = (\bar{q}_1(\omega), \bar{q}_2(\omega)) \in \partial h_2(\bar{x}_1, \bar{x}_2(\omega); \omega) \quad \text{a.s.} \quad (23)$$

Let  $\iota_1 : X_1 \rightarrow L^\infty(\Omega, X_1)$  be the continuous injection such that  $\iota(x) = (\iota_1 x_1, x_2)$ . Notice that for  $\bar{q}_1^* \in L^1(\Omega, X_1^*)$ , the adjoint  $\iota_1^* : (L^\infty(\Omega, X_1))^* \rightarrow X_1^*$  satisfies

$$\langle \iota_1^* \bar{q}_1^*, x_1 \rangle_{X_1^*, X_1} = \langle \bar{q}_1, \iota_1 x_1 \rangle_{L^1(\Omega, X_1^*), L^\infty(\Omega, X_1)} = \mathbb{E}[\langle \bar{q}_1(\cdot), x_1 \rangle_{X_1^*, X_1}] \quad \forall x_1 \in X_1.$$

Hence  $\iota^* \bar{q} = (\mathbb{E}[\bar{q}_1], \bar{q}_2) \in X_1^* \times L^1(\Omega, X_2^*)$ . Thus  $-\iota^* \bar{q} \in \partial H_1(\bar{x})$  can be equivalently written as

$$H_1(x) \geq H_1(\bar{x}) - \langle \mathbb{E}[\bar{q}_1], x_1 - \bar{x}_1 \rangle_{X_1^*, X_1} - \mathbb{E}[\langle \bar{q}_2, x_2 - \bar{x}_2 \rangle_{X_2^*, X_2}]$$

for all  $x \in X$ . We obtain from (20)

$$J_1(x_1) + \ell(x_1, \bar{\lambda}^\circ) \geq J_1(\bar{x}_1) + \ell(\bar{x}_1, \bar{\lambda}^\circ) - \langle \mathbb{E}[\bar{q}_1], x_1 - \bar{x}_1 \rangle_{X_1^*, X_1} \quad \forall x_1 \in C \quad (24)$$

and

$$\mathbb{E}[\langle \bar{q}_2, x_2 - \bar{x}_2 \rangle_{X_2^*, X_2}] \geq 0 \quad \forall x_2 \in L^\infty(\Omega, X_2). \quad (25)$$

The expression (24) is clearly equivalent to condition (i). Moreover, (25) implies  $\bar{q}_2(\omega) = 0$  a.s. In combination with (23), this implies that for all  $(x_1, x_2) \in X_1 \times X_2$ ,

$$h_2(x_1, x_2; \omega) \geq h_2(\bar{x}_1, \bar{x}_2(\omega); \omega) + \langle \bar{q}_1(\omega), x_1 - \bar{x}_1 \rangle_{X_1^*, X_1} \quad \text{a.s.}$$

From the definition of  $h_2$ , it follows that

$$\begin{aligned} & J_2(x_1, x_2; \omega) \bar{\vartheta}(\omega) + \langle \bar{\lambda}_e(\omega), e(x_1, x_2; \omega) \rangle_{W^*, W} \\ & + \langle \bar{\lambda}_i(\omega), i(x_1, x_2; \omega) \rangle_{R^*, R} - \langle \bar{q}_1(\omega), x_1 \rangle_{X_1^*, X_1} \\ & \geq J_2(\bar{x}_1, \bar{x}_2(\omega); \omega) \bar{\vartheta}(\omega) + \langle \bar{\lambda}_e(\omega), e(\bar{x}_1, \bar{x}_2(\omega); \omega) \rangle_{W^*, W} \\ & + \langle \bar{\lambda}_i(\omega), i(\bar{x}_1, \bar{x}_2(\omega); \omega) \rangle_{R^*, R} - \langle \bar{q}_1(\omega), \bar{x}_1 \rangle_{X_1^*, X_1} \end{aligned} \quad (26)$$

for all  $(x_1, x_2) \in X_1 \times X_2$ . The inequality (26) is clearly equivalent to condition (ii) with  $\bar{\rho}(\omega) := \bar{q}_1(\omega)$ .  $\square$

**Remark 3.11.** Implicitly proven in Theorem 3.9 are some additional properties that will later be convenient. First,  $f(x_1, x_2; \omega) := J_2(x_1, x_2; \omega) \bar{\vartheta}(\omega)$  is Carathéodory and convex with respect to  $(x_1, x_2)$ . The growth conditions ensure that  $F(\chi) = \mathbb{E}[f(\chi_1(\cdot), \chi_2(\cdot); \cdot)]$  is continuous with respect to  $\chi \in \mathfrak{X} = L^\infty(\Omega, X_1) \times L^\infty(\Omega, X_2)$ . Therefore, an element of the subdifferential  $\partial F(\iota(\bar{x}))$  can be identified with a pair  $(\bar{\phi}_1, \bar{\phi}_2)$  belonging to a weakly compact subset in  $L^1(\Omega, X_1^*) \times L^1(\Omega, X_2^*)$  such that

$$\bar{\phi}(\omega) = (\bar{\phi}_1(\omega), \bar{\phi}_2(\omega)) \in \partial f(\bar{x}_1, \bar{x}_2(\omega); \omega) = \partial J_2(\bar{x}_1, \bar{x}_2(\omega); \omega) \bar{\vartheta}(\omega) \quad \text{a.s.}$$

The equality of the subdifferentials follows since  $\bar{\vartheta} \geq 0$  a.s.

Additionally, we show that  $\bar{\zeta}$  satisfying  $\bar{\zeta}(\omega) \in \partial J_2(\bar{x}_1, \bar{x}_2(\omega); \omega)$  a.s. belongs to the set  $L^p(\Omega, X_1^*) \times L^p(\Omega, X_2^*)$ . By definition of the subdifferential, we have for all  $\tilde{\chi} \in \mathfrak{X}$

$$\langle \bar{\zeta}(\omega), \tilde{\chi}(\omega) - \bar{x}(\omega) \rangle_{X_1^* \times X_2^*, X_1 \times X_2} \leq J_2(\tilde{\chi}_1(\omega), \tilde{\chi}_2(\omega); \omega) - J_2(\bar{x}_1, \bar{x}_2(\omega); \omega) \quad \text{a.s.}$$

Hence with  $\chi := \tilde{\chi} - \bar{x}$  we obtain  $\langle \bar{\zeta}(\omega), \chi(\omega) \rangle_{X_1^* \times X_2^*, X_1 \times X_2} \leq J_2(\chi_1(\omega), \chi_2(\omega); \omega)$  a.s. For any  $r > 0$ , there exists by Assumption 3.1(iii) a function  $a_r \in L^p(\Omega)$  such that  $|J_2(\chi_1(\omega), \chi_2(\omega); \omega)| \leq a_r(\omega)$  for  $\chi$  satisfying  $\|\chi\|_{\mathfrak{X}} < r$ . Therefore

$$\|\bar{\zeta}(\omega)\|_{X_1^* \times X_2^*} = \sup_{\chi(\omega): \|\chi(\omega)\|_{X_1 \times X_2} \leq 1} \langle \bar{\zeta}(\omega), \chi(\omega) \rangle_{X_1^* \times X_2^*, X_1 \times X_2} \leq a_r(\omega),$$

so it follows that  $\mathbb{E}[\|\bar{\zeta}_i\|_{X_i^*}^p] < \infty$  for  $i = 1, 2$ . Since  $\bar{\zeta}$  is measurable, it follows that  $\bar{\zeta} \in L^p(\Omega, X_1^*) \times L^p(\Omega, X_2^*)$  as claimed.

### 3.4 Existence of regular multipliers

In this section, we refine the results by [19] to include objective functions with risk measures. The additional Assumption 3.6 ensures that Lagrange multipliers are more regular.

**Lemma 3.12.** Let Assumption 3.1, Assumption 3.5, and Assumption 3.6 be satisfied. Then there exists a saddle point  $(\bar{x}, \bar{\lambda}) \in X_0 \times \Lambda_0$  to the Lagrangian  $L$ , i.e.,  $(\bar{x}, \bar{\lambda})$  satisfies

$$L(\bar{x}, \lambda) \leq L(\bar{x}, \bar{\lambda}) \leq L(x, \bar{\lambda}) \quad \forall (x, \lambda) \in X \times \Lambda. \quad (27)$$

*Proof.* This is proven in Theorem B.2 and Theorem B.4. □

**Theorem 3.13.** Let Assumption 3.1 be satisfied. Then  $(\bar{x}, \bar{\lambda}) \in X_0 \times \Lambda_0$  satisfies (27) if and only if there exist  $\bar{\rho} \in L^1(\Omega, X^*)$  and  $\bar{\vartheta} \in \partial \mathcal{R}[J_2(\bar{x})]$  fulfilling the following conditions:

(i) The function

$$x_1 \mapsto J_1(x_1) + \langle \mathbb{E}[\bar{\rho}], x_1 \rangle_{X_1^*, X_1}$$

attains its minimum over  $C$  at  $\bar{x}_1$ .

(ii) The function

$$\begin{aligned} (x_1, x_2) \mapsto & J_2(x_1, x_2; \omega) \bar{\vartheta}(\omega) + \langle \bar{\lambda}_e(\omega), e(x_1, x_2; \omega) \rangle_{W^*, W} \\ & + \langle \bar{\lambda}_i(\omega), i(x_1, x_2; \omega) \rangle_{R^*, R} - \langle \bar{\rho}(\omega), x_1 \rangle_{X_1^*, X_1} \end{aligned}$$

attains its minimum in  $X_1 \times X_2$  at  $(\bar{x}_1, \bar{x}_2(\omega))$  for almost every  $\omega \in \Omega$ .

(iii) It holds that  $\bar{x}_1 \in C$  and the following conditions hold almost surely:

$$\begin{aligned} e(\bar{x}_1, \bar{x}_2(\omega); \omega) &= 0, \\ \bar{\lambda}_i(\omega) &\in K^\oplus, \quad i(\bar{x}_1, \bar{x}_2(\omega); \omega) \leq_K 0, \quad \langle \bar{\lambda}_i(\omega), i(\bar{x}_1, \bar{x}_2(\omega); \omega) \rangle_{R^*, R} = 0. \end{aligned}$$

*Proof.* The proof follows the arguments from Theorem 3.9 after setting  $L^\circ \equiv 0$ .  $\square$

In the risk-neutral case  $\mathcal{R} = \mathbb{E}$ , the subdifferential consists of the unique subgradient  $\bar{\vartheta} \equiv 1$ . In that case, Theorem 3.13 recovers the optimality conditions presented in [19].

## 4 Moreau–Yosida Regularization

Optimal control problems with state constraints were first handled in [11] and the use of a Moreau–Yosida regularization for such problems was first studied in [21, 22]. In this section, we detail the use of Moreau–Yosida regularization for the state constraint

$$i(x_1, x_2(\omega); \omega) \leq_K 0 \quad (\Leftrightarrow -i(x_1, x_2(\omega); \omega) \in K) \quad (28)$$

from Problem (P), i.e., in the novel setting of stochastic optimization problem with almost sure constraints.

### 4.1 Moreau–Yosida Revisited

In applications, it is often the case that  $K$  coincides with a cone on a weaker space than  $R$ , the space where  $K$  has a nonempty interior. Let  $(H, (\cdot, \cdot)_H)$  be a Hilbert space. We consider cases in which the image space  $R$  of the conical constraint belongs to a Gelfand triple in the sense that

$$R \hookrightarrow H \cong H^* \hookrightarrow R^*, \quad (29)$$

where  $\hookrightarrow$  denotes the dense and continuous embedding. In particular  $\langle h, r \rangle_{R^*, R} = (h, r)_H$  whenever  $r \in R$  and  $h \in H$ . We assume the cone  $K$  is compatible with a cone  $K_H$  on the less regular space  $H$  in the sense that  $K_H \cap R = K$ . This construction is quite natural and allows the construction of a regularization with respect to the weaker  $H$ -norm, where computations are cheaper. Notably, since  $i$  has an image in  $R$ , it follows that

$$i(x_1, x_2(\omega); \omega) \leq_{K_H} 0 \quad \Leftrightarrow \quad i(x_1, x_2(\omega); \omega) \leq_K 0. \quad (30)$$

Examples of applications are given at the end of this section.

Given a parameter  $\gamma > 0$ , the Moreau–Yosida regularization (or envelope) for  $\delta_{K_H}$  is given by the function  $\beta^\gamma : H \rightarrow \mathbb{R}$  with

$$\beta^\gamma(k) = \inf_{y \in H} \left\{ \delta_{K_H}(y) + \frac{\gamma}{2} \|k - y\|_H^2 \right\}.$$

Clearly,  $\beta^\gamma$  has the property that  $\beta^\gamma(k) = 0$  whenever  $k \in K_H$ , and otherwise  $\beta^\gamma(k) > 0$ . Additionally,

$$\beta^\gamma(k) = \frac{\gamma}{2} \inf_{y \in H} \left\{ \delta_{K_H}(y) + \|k - y\|_H^2 \right\} = \frac{\gamma}{2} \|k - \pi_{K_H}(k)\|_H^2, \quad (31)$$

where  $\pi_{K_H} : H \rightarrow K_H, k \mapsto \arg \min_{y \in H} \{\delta_{K_H}(y) + \|k - y\|_H\}$  denotes the projection onto  $K_H$ . Let  $K_H^\oplus = \{h \in H : (h, k)_H \geq 0 \forall k \in K_H\}$  denote the dual cone to  $K$ . The projection  $\pi_{K_H}$  onto a nonempty, closed, convex cone  $K_H$  admits the characterization (cf. [7, Proposition 6.27])

$$\pi_{K_H}(x) = p \iff p \in K_H, (p, x - p)_H = 0, \text{ and } p - x \in K_H^\oplus. \quad (32)$$

Moreover (cf. [7, Proposition 12.29]),  $\beta^\gamma$  is Fréchet differentiable on  $H$  with the gradient  $\nabla \beta^\gamma = \gamma(\text{id}_H - \pi_{K_H})$ .

The following result is proven in Appendix C and is needed in the following Lemma 4.2.

**Lemma 4.1.** Let  $Y$  and  $Z$  be real Banach spaces and  $K \subset Z$  be a nonempty, closed, and convex cone. Suppose  $f : Y \rightarrow Z$  is  $K$ -convex and  $g : Z \rightarrow \mathbb{R}$  is convex having the property that  $g(z) = 0$  if  $z \in K$ . Then  $g \circ (-f)$  is convex.

**Lemma 4.2.** Under Assumption 3.1,  $x \mapsto \mathbb{E}[\beta^\gamma(-i(x))]$  is weakly\* lower semicontinuous on  $X$  for any  $\gamma > 0$ . Moreover, we have  $\mathbb{E}[\beta^\gamma(-i(x))] = 0$  if and only if  $i(x_1, x_2(\omega); \omega) \leq_K 0$  a.s.

*Proof.* First, we prove that  $h(x_1, x_2; \omega) := \beta^\gamma(-i(x_1, x_2; \omega))$  is a normal convex integrand. Since the indicator function  $\delta_{K_H}$  is a proper, convex, and lower semicontinuous function on  $H$ ,  $\beta^\gamma$  is a convex, real-valued, and continuous function (cf. [7, Proposition 12.15]). In particular,  $\beta^\gamma$  is a normal integrand, so  $\omega \mapsto \beta^\gamma(f(\omega))$  is measurable for any measurable mapping  $f : \Omega \rightarrow \mathbb{R}$ . Continuity of both  $\beta^\gamma$  and  $i$  imply continuity of  $h$  on  $X_1 \times X_2$ ; convexity of  $h$  on  $X_1 \times X_2$  follows by Lemma 4.1. As a result,  $h$  is a normal convex integrand.

Let  $r > 0$  and suppose  $(x_1, x_2) \in X_1 \times X_2$  satisfies  $\|x_1\|_{X_1} + \|x_2\|_{X_2} \leq r$ . Since  $0 \in K = K_H \cap R$  and  $\pi_{K_H}$  is nonexpansive, we have thanks to the characterization (31)

$$\begin{aligned} h(x_1, x_2; \omega) &= \frac{\gamma}{2} \|-i(x_1, x_2; \omega) - \pi_{K_H}(-i(x_1, x_2; \omega))\|_H^2 \\ &\leq \gamma(\|-i(x_1, x_2; \omega) - 0\|_H^2 + \|\pi_{K_H}(0) - \pi_{K_H}(-i(x_1, x_2; \omega))\|_H^2) \\ &\leq 2\gamma\|i(x_1, x_2; \omega)\|_H^2 \leq 2\gamma cb_{r,i}^2, \end{aligned}$$

where  $b_{r,i}$  is defined as in Assumption 3.1(v) and  $c > 0$  depends on the embedding constant from  $R \hookrightarrow H$ . Hence  $h(x_1, x_2(\omega); \omega)$  is integrable in  $\omega$  for all  $x$ . Weak\* lower semicontinuity of  $I_h$  follows by Remark 3.3. The second statement is clear when considering (30) in combination with  $\beta^\gamma \geq 0$  and  $\beta^\gamma(k) = 0$  if and only if  $k \in K_H$ .  $\square$

In the introduction, we already gave one example ( $P'$ ) of an application to PDE-constrained optimization under uncertainty. Now, we will consider three representative examples of the almost sure constraint (28) from the same problem class. Additionally, we provide their regularization functions.

**Example 4.3** (Mixed control/state constraints). Assume  $X_2 = H^1(D)$ ,  $X_1 = H = L^2(D)$ , and  $K_H = \{z \in H : z(s) \geq 0 \text{ a.e. in } D\}$ . For  $a \in L^\infty(\Omega, X_2)$ , the pointwise almost sure constraint

$$x_2(s, \omega) \leq a(s, \omega) + \varepsilon x_1(s), \quad \text{a.e. } s \in D, \text{ a.s. } \omega \in \Omega$$

has the form (28) with  $i(x_1(\cdot), x_2(\cdot, \omega); \omega) = x_2(\cdot, \omega) - a(\cdot, \omega) - \varepsilon x_1(\cdot)$ . In the case  $\varepsilon = 0$ ,  $R = H^1(D)$  and we have a stochastic version of state constraints originally investigated by [11]. In the case  $\varepsilon > 0$ ,  $R = L^2(D)$ , and we have a stochastic version of a bottleneck constraint as investigated by [9].

By (31),

$$\beta^\gamma(-i(x_1(\cdot), x_2(\cdot, \omega); \omega)) = \frac{\gamma}{2} \|\max(0, x_2(\cdot, \omega) - a(\cdot, \omega) - \varepsilon x_1(\cdot))\|_{L^2(D)}^2,$$

where  $\max$  is defined pointwise in  $D$ .

**Example 4.4** (Volume constraint on state). Assume  $X_2 = H^1(D)$ ,  $H = R = \mathbb{R}$ , and  $K_H = \mathbb{R}_+$ . For  $b \in L^\infty(\Omega)$ , the constraint

$$\int_D x_2(s, \omega) \, ds \leq b(\omega), \quad \text{a.s. } \omega \in \Omega$$

takes the form  $i(x_2(\cdot, \omega); \omega) = \int_D x_2(s, \omega) \, ds - b(\omega)$  and the regularization function is given by

$$\beta^\gamma(-i(x_2(\cdot, \omega); \omega)) = \frac{\gamma}{2} \left| \max \left( 0, \int_D x_2(s, \omega) \, ds - b(\omega) \right) \right|^2.$$

**Example 4.5** (State gradient constraint). Assume  $X_2 = H^1(D)$ ,  $H = R = L^2(D)$ , and  $K_H$  is the same cone as in Example 4.3. For  $\psi \in L^\infty(\Omega, H)$ , the almost sure constraint

$$|\nabla x_2(s, \omega)|_2 \leq \psi(s, \omega) \quad \text{a.e. } s \in D, \text{ a.s. } \omega \in \Omega$$

takes the form  $i(x_2(\cdot, \omega); \omega) = |\nabla x_2(\cdot, \omega)|_2 - \psi(\cdot, \omega)$ . The regularization is

$$\beta^\gamma(-i(x_2(\cdot, \omega); \omega)) = \frac{\gamma}{2} \|\max(0, |\nabla x_2(\cdot, \omega)|_2 - \psi(\cdot, \omega))\|_{L^2(D)}^2.$$

## 4.2 Convergence of Regularized Problems

We now discuss the Moreau–Yosida regularization of the conical constraint from Problem (P). To that end, we define the family of regularized problems, parametrized by  $\gamma > 0$ , by

$$\begin{aligned} & \underset{x \in X}{\text{minimize}} \quad \{j^\gamma(x) := j(x) + \mathbb{E}[\beta^\gamma(-i(x))]\} \\ & \text{s.t.} \quad \begin{cases} x_1 \in C, \\ e(x_1, x_2(\omega); \omega) = 0 \quad \text{a.s.} \end{cases} \end{aligned} \quad (\text{P}^\gamma)$$

Problem (P $^\gamma$ ) is interesting in its own right. Due to the linearity of the constraint, in the case where  $j$  is quadratic, it is related to a subproblem used in sequential quadratic programming.

Thanks to Assumption 3.1, the functions  $e(\cdot, \cdot; \omega)$  are Fréchet differentiable with partial derivatives that are bounded and linear, i.e.,  $e_{x_i}(\omega) \in \mathcal{L}(X_i, W)$ . Without loss of generality, affine terms can be incorporated in the objective function, so by linearity,

$$e(x_1, x_2; \omega) = e_{x_1}(\omega)x_1 + e_{x_2}(\omega)x_2.$$

We focus on the case where Problem (P) does not satisfy the relatively complete recourse condition. This occurs naturally when the equality constraint has a unique solution for every  $x_1 \in C$ . Then it is clear that the relatively complete recourse condition will not be satisfied except for trivial choices of  $K$ .

For the next results, we need the following additional assumptions.

**Assumption 4.6.** (i)  $e_{x_2}(\omega)$  is a linear isomorphism for almost every  $\omega$  and  $e_{x_2}^{-1} \in L^\infty(\Omega, \mathcal{L}(W, X_2))$ .  
(ii)  $i$  is continuously Fréchet differentiable on  $X_1 \times X_2$  for almost every  $\omega$  with partial derivatives satisfying  $i_{x_k}(x_1, x_2; \cdot) \in L^\infty(\Omega, \mathcal{L}(X_k, R))$  for every  $(x_1, x_2) \in X_1 \times X_2$  and  $k = 1, 2$ .  
(iii) If  $x^\gamma \rightharpoonup^* \bar{x}$  and  $j(x^\gamma) \rightarrow j(\bar{x})$ , then  $\|x_1^\gamma\|_{X_1} \rightarrow \|\bar{x}_1\|_{X_1}$ .

**Remark 4.7.** For the consistency of the primal problems, shown in Proposition 4.8, we only need Assumption 4.6(i). By Assumption 4.6(i), for every  $x_1 \in C$ , there exists a unique solution  $x_2 \in L^\infty(\Omega, X_2)$  such that

$$x_2(\omega) = -e_{x_2}^{-1}(\omega)e_{x_1}(\omega)x_1. \quad (33)$$

Thanks to the additional structure afforded by Assumption 4.6(i), it is in fact enough to require that  $j$  is radially unbounded with respect to  $x_1$  only (instead of  $x$ ). Assumption 4.6(iii) is a property that is satisfied in combination with (33), e.g., if  $J_1$  is a  $\mu$ -strongly convex function, as is the case for the Tichonov regularization  $J_1(x_1) = \alpha\|x_1\|_{X_1}^2$ ,  $\alpha > 0$ . To see this, notice that by (33), the control-to-state operator  $S(\omega) : X_1 \rightarrow X_2$ ,  $x_1 \mapsto x_2(\omega)$  is a linear mapping for almost every  $\omega$ . The reduced functional  $\hat{J}_2(x_1) := \mathcal{R}[J_2(x_1, S(\cdot)x_1; \cdot)]$  is therefore convex. It is straightforward to prove that  $\hat{j}(x_1^\gamma) := J_1(x_1^\gamma) + \hat{J}_2(x_1^\gamma) = j(x^\gamma)$  is a  $\mu$ -strongly convex function over  $X_1$ . Note that if  $x^\gamma \rightharpoonup^* \bar{x}$ , then  $x_1^\gamma \rightharpoonup^* \bar{x}_1$  and thus the convex combination  $z_t^\gamma := tx_1^\gamma + (1-t)\bar{x}_1 \in C$  satisfies  $z_t^\gamma \rightharpoonup^* \bar{x}_1$  for any  $t \in [0, 1]$ . By weak\* lower semicontinuity and convexity of  $\hat{j}$  as well as  $j(x^\gamma) \rightarrow j(\bar{x})$ ,

$$\hat{j}(\bar{x}_1) \leq \liminf_{\gamma \rightarrow \infty} \hat{j}(z_t^\gamma) \leq \limsup_{\gamma \rightarrow \infty} \hat{j}(z_t^\gamma) \leq \limsup_{\gamma \rightarrow \infty} t\hat{j}(x_1^\gamma) + (1-t)\hat{j}(\bar{x}_1) = \hat{j}(\bar{x}_1),$$

meaning  $\lim_{\gamma \rightarrow \infty} \hat{j}(tx_1^\gamma + (1-t)\bar{x}_1) = \hat{j}(\bar{x}_1)$  for any  $t \in [0, 1]$ . Now,  $\mu$ -strong convexity of  $\hat{j}$  gives

$$\begin{aligned} \lim_{\gamma \rightarrow \infty} \mu t(1-t)\|x_1^\gamma - \bar{x}_1\|_{X_1}^2 + \lim_{\gamma \rightarrow \infty} \hat{j}(tx_1^\gamma + (1-t)\bar{x}_1) &\leq \lim_{\gamma \rightarrow \infty} t\hat{j}(x_1^\gamma) + (1-t)\hat{j}(\bar{x}_1) \\ \Leftrightarrow \lim_{\gamma \rightarrow \infty} \mu t(1-t)\|x_1^\gamma - \bar{x}_1\|_{X_1}^2 &\leq 0, \end{aligned}$$

which gives  $\|x_1^\gamma\|_{X_1} \rightarrow \|\bar{x}_1\|_{X_1}$  as required.

The first result concerns optimality of limit points of solutions to Problem  $(P^\gamma)$ .

**Proposition 4.8.** Suppose Problem  $(P)$  satisfies Assumption 3.1, Assumption 3.5, and Assumption 4.6(i). Then there exists a solution  $x^\gamma \in X$  to Problem  $(P^\gamma)$  for all  $\gamma > 0$ . Additionally, given  $\{\gamma_n\}$  with  $\gamma_n \rightarrow \infty$ , weak\* limit points of  $\{x^{\gamma_n}\}$  solve Problem  $(P)$ .

*Proof.* Let  $F'_{\text{ad}} := \{x \in X : x_1 \in C, e(x_1, x_2(\omega); \omega) = 0 \text{ a.s.}\}$  denote the feasible set for Problem  $(P^\gamma)$ . From (33), we have

$$\|x_2\|_{L^\infty(\Omega, X_2)} \leq \|e_{x_2}^{-1}\|_{L^\infty(\Omega, \mathcal{L}(W, X_2))} \|e_{x_1} x_1\|_{L^\infty(\Omega, W)} \leq c \|x_1\|_{X_1}. \quad (34)$$

Either boundedness of  $C$  or radial unboundedness of  $j$  imply boundedness of  $j^\gamma$  over  $F'_{\text{ad}}$ . Thus a minimizing sequence  $\{x^n\} \subset F'_{\text{ad}}$  exists such that

$$\lim_{n \rightarrow \infty} j^\gamma(x^n) = \bar{j}^\gamma := \inf_{x \in F'_{\text{ad}}} j^\gamma(x). \quad (35)$$

If  $C$  is bounded, then (34) implies  $x^n$  is bounded in  $X$  for all  $n$ . Otherwise, if  $j$  is radially unbounded, then for every  $c > 0$  there exists  $r > 0$  such that for all  $n$  satisfying  $\|x^n\|_X > r$ , we have  $j^\gamma(x^n) \geq j(x^n) > c$ . Choosing  $c$  large enough produces a contradiction to (35). Hence  $x^n$  is bounded for all  $n$ , so there exists a subsequence  $\{n_k\}$  such that  $x^{n_k} \rightharpoonup^* x^\gamma$ . Assumption 3.1 guarantees that

Proposition 3.4 and Lemma 4.2 apply, from which we deduce weak\* lower semicontinuity of  $j^\gamma$ . As a result, we obtain

$$\bar{j}^\gamma \leq j^\gamma(x^\gamma) \leq \lim_{k \rightarrow \infty} j^\gamma(x^{n_k}) = \bar{j}^\gamma.$$

Notice that  $C$  is closed and by Corollary 2.7,  $F'_{\text{ad}}$  is weakly\* closed. Hence  $x^\gamma \in F'_{\text{ad}}$  and therefore  $x^\gamma$  is an optimum for Problem (P $^\gamma$ ).

Now, observe a sequence  $\{x^{\gamma_n}\}$  of optima to Problem (P $^\gamma$ ) with  $\gamma_n \rightarrow \infty$ . If  $\bar{x}$  is an optimum of Problem (P), then by optimality,

$$j(x^{\gamma_n}) \leq j^{\gamma_n}(x^{\gamma_n}) \leq j^{\gamma_n}(\bar{x}) = j(\bar{x}). \quad (36)$$

In fact (36) implies that  $\{x^{\gamma_n}\}$  makes up a minimizing sequence for  $j$ . Again, boundedness of  $C$  or radial unboundedness of  $j$  imply boundedness of  $\{x^{\gamma_n}\}$ . Therefore, there exists a subsequence  $\{n_k\}$  such that  $x^{\gamma_{n_k}} \rightharpoonup^* \hat{x}$ . Weak\* lower semicontinuity of  $j$  gives  $j(\hat{x}) \leq j(\bar{x})$ . Now, it is enough to show feasibility of  $\hat{x}$ , since from feasibility one must conclude  $j(\hat{x}) = j(\bar{x})$  by optimality of  $\bar{x}$ . Notice that  $j$  is bounded over  $F'_{\text{ad}}$ , so from (36), we have

$$\mathbb{E}[\beta^{\gamma_{n_k}}(-i(x^{\gamma_{n_k}}))] = \frac{\gamma_{n_k}}{2} \mathbb{E}[\| -i(x^{\gamma_{n_k}}) - \pi_{K_H}(-i(x^{\gamma_{n_k}})) \|_H^2] \leq c.$$

Using the weak\* lower semicontinuity granted by Lemma 4.2, we obtain the convergence  $\mathbb{E}[\| -i(x^{\gamma_{n_k}}) - \pi_{K_H}(-i(x^{\gamma_{n_k}})) \|_H^2] \rightarrow 0$  as  $k \rightarrow \infty$ , meaning  $-i(\hat{x}_1, \hat{x}_2(\omega); \omega) \in K$  a.s. by Lemma 4.2. Feasibility of  $\hat{x}$  for Problem (P) now follows by closedness of  $C$  and Corollary 2.7. In closing, we have  $j(x^{\gamma_n}) \rightarrow j(\hat{x}) = j(\bar{x})$  since

$$j(\hat{x}) \leq \liminf_{\gamma_n \rightarrow \infty} j(x^{\gamma_n}) \leq \limsup_{\gamma_n \rightarrow \infty} j(x^{\gamma_n}) \leq j(\bar{x}) = j(\hat{x}). \quad (37)$$

□

In the following, we analyze the convergence of Lagrange multipliers as  $\gamma \rightarrow \infty$ . To that end, it will be helpful to first formulate more explicit optimality conditions for Problem (P), and after that the explicit conditions for Problem (P $^\gamma$ ). In what follows, we use the abusive notation  $e_{x_i}^*$  for adjoint operator from  $(L^\infty(\Omega, W))^*$  to  $(L^\infty(\Omega, X_i))^*$  and  $e_{x_i}^*(\cdot)$  for the adjoint operator from  $L^1(\Omega, W^*)$  to  $L^1(\Omega, X_i^*)$  as described in Lemma 2.6. Similarly,  $i_{x_i}^*(\bar{x})$  is the shorthand for the operator from  $(L^\infty(\Omega, R))^*$  to  $(L^\infty(\Omega, X_i))^*$  whereas  $i_{x_i}^*(\bar{x}_1, \bar{x}_2(\cdot); \cdot)$  represents the operator from  $L^1(\Omega, R^*)$  to  $L^1(\Omega, X_i^*)$ .

**Lemma 4.9.** Suppose Assumption 3.1, Assumption 3.5, and Assumption 4.6(i)–4.6(ii) are satisfied. Then there exists a saddle point  $(\bar{x}, \bar{\lambda}, \bar{\lambda}^\circ)$  of the Lagrangian  $\bar{L}$ . Furthermore, the existence of the saddle point is equivalent to the following necessary and sufficient optimality conditions: there exist  $\bar{\rho} \in L^1(\Omega, X_1^*)$ ,  $\bar{\vartheta} \in \partial \mathcal{R}[J_2(\bar{x})]$ ,  $\bar{\eta} \in J_1(\bar{x}_1)$ ,  $\bar{\xi} \in N_C(\bar{x}_1)$ , and  $\bar{\zeta} = (\bar{\zeta}_1, \bar{\zeta}_2) \in L^p(\Omega, X_1^*) \times L^p(\Omega, X_2^*)$  with  $\bar{\zeta}(\omega) \in \partial J_2(\bar{x}_1, \bar{x}_2(\omega); \omega)$  such that

$$\bar{\eta} + \mathbb{E}[\bar{\rho}] + e_{x_1}^* \bar{\lambda}_e^\circ + i_{x_1}^*(\bar{x}) \bar{\lambda}_i^\circ + \bar{\xi} = 0, \quad (38a)$$

$$e_{x_2}^* \bar{\lambda}_e^\circ + i_{x_2}^*(\bar{x}) \bar{\lambda}_i^\circ = 0, \quad (38b)$$

$$\bar{\zeta}_1(\omega) \bar{\vartheta}(\omega) + e_{x_1}^*(\omega) \bar{\lambda}_e(\omega) + i_{x_1}^*(\bar{x}_1, \bar{x}_2(\omega); \omega) \bar{\lambda}_i(\omega) - \bar{\rho}(\omega) = 0, \quad (38c)$$

$$\bar{\zeta}_2(\omega) \bar{\vartheta}(\omega) + e_{x_2}^*(\omega) \bar{\lambda}_e(\omega) + i_{x_2}^*(\bar{x}_1, \bar{x}_2(\omega); \omega) \bar{\lambda}_i(\omega) = 0, \quad (38d)$$

$$\bar{x}_1 \in C, \quad e(\bar{x}_1, \bar{x}_2(\omega); \omega) = 0, \quad (38e)$$

$$i(\bar{x}_1, \bar{x}_2(\omega); \omega) \leq_K 0, \quad \bar{\lambda}_i(\omega) \in K^\oplus, \quad \langle \bar{\lambda}_i(\omega), i(\bar{x}_1, \bar{x}_2(\omega); \omega) \rangle_{R^*, R} = 0, \quad (38f)$$

$$\bar{\lambda}_i^\circ \in \mathcal{K}^\oplus, \quad \langle \bar{\lambda}_i^\circ, i(\bar{x}) \rangle_{(L^\infty(\Omega, R))^*, L^\infty(\Omega, R)} = 0, \quad (38g)$$

with pointwise conditions holding for almost all  $\omega \in \Omega$ .

*Proof.* By Lemma 3.12, a saddle point  $(\bar{x}, \bar{\lambda}, \bar{\lambda}^\circ)$  exists, and so we can apply Theorem 3.9. The conditions (38e)–(38g) are trivially satisfied. Minimizing  $J_1(x_1) + \langle \mathbb{E}[\bar{\rho}], x_1 \rangle_{X_1^*, X_1} + \ell(x_1, \bar{\lambda}^\circ)$  in  $x_1$  over  $C$  is equivalent to minimizing

$$F(x) := J_1(x_1) + \langle \mathbb{E}[\bar{\rho}], x_1 \rangle_{X_1^*, X_1} + \langle \bar{\lambda}_e^\circ, e(x) \rangle_{(L^\infty(\Omega, W))^*, L^\infty(\Omega, W)} \\ + \langle \bar{\lambda}_i^\circ, i(x) \rangle_{(L^\infty(\Omega, R))^*, L^\infty(\Omega, R)} + \delta_C(x_1)$$

over  $X$ . In combination with (38e)–(38f), the condition  $\ell(\bar{x}_1, \bar{\lambda}^\circ) = 0$  implies the minimum in the definition of (12) can only be attained at  $\bar{x}_2$  by Remark 3.10. Hence condition (i) of Theorem 3.9 is the same as  $0 \in \partial F(\bar{x})$ . All terms in  $F$  other than  $\delta_C$  are continuous on  $X$ . Thus the sum rule for the subdifferential applies (cf. [10, Theorem 2.168]) and we have after exploiting differentiability afforded by Assumption 4.6(ii) the expression

$$0 \in \begin{pmatrix} \partial J_1(\bar{x}_1) \\ \{0\} \end{pmatrix} + \begin{pmatrix} \{\mathbb{E}[\bar{\rho}] + e_{x_1}^* \bar{\lambda}_e^\circ + i_{x_1}^*(\bar{x}) \bar{\lambda}_i^\circ\} \\ \{e_{x_2}^* \bar{\lambda}_e^\circ + i_{x_2}^*(\bar{x}) \bar{\lambda}_i^\circ\} \end{pmatrix} + \begin{pmatrix} \partial \delta_C(\bar{x}_1) \\ \{0\} \end{pmatrix}.$$

Note that the corresponding adjoint operators are well-defined by Lemma 2.6. The existence of  $\bar{\eta}$  and  $\bar{\xi}$  as well as (38a)–(38b) follow. Condition (ii) of Theorem 3.9 yields

$$0 \in \partial(J_2(\bar{x}_1, \bar{x}_2(\omega); \omega) \bar{\vartheta}(\omega)) + \begin{pmatrix} \{e_{x_1}^*(\omega) \bar{\lambda}_e(\omega) + i_{x_1}^*(\bar{x}_1, \bar{x}_2(\omega); \omega) \bar{\lambda}_i(\omega) - \bar{\rho}(\omega)\} \\ \{e_{x_2}^*(\omega) \bar{\lambda}_e(\omega) + i_{x_2}^*(\bar{x}_1, \bar{x}_2(\omega); \omega) \bar{\lambda}_i(\omega)\} \end{pmatrix}.$$

Since  $\bar{\vartheta} \geq 0$  a.s.,  $\partial(J_2(\bar{x}_1, \bar{x}_2(\omega); \omega) \bar{\vartheta}(\omega)) = \partial J_2(\bar{x}_1, \bar{x}_2(\omega); \omega) \bar{\vartheta}(\omega)$ . Hence there exists  $\bar{\zeta}(\omega) \in \partial J_2(\bar{x}_1, \bar{x}_2(\omega); \omega)$  such that (38c)–(38d) are satisfied a.s. The fact that  $\bar{\zeta} \in L^p(\Omega, X_1^*) \times L^p(\Omega, X_2^*)$  follows from Remark 3.11.  $\square$

Now, we formulate the optimality conditions for Problem  $(P^\gamma)$ . We define the corresponding Lagrangian by  $L^\gamma(x, \lambda) := j^\gamma(x) + \mathbb{E}[\langle \lambda_e, e(x) \rangle_{W^*, W}]$ .

**Lemma 4.10.** Suppose Assumption 3.1, Assumption 3.5, and Assumption 4.6(i)–4.6(ii) are satisfied. Then there exists a saddle point  $(x^\gamma, \lambda^\gamma)$  of the Lagrangian  $L^\gamma$ . Furthermore, the existence of the saddle point is equivalent to the following necessary and sufficient optimality conditions: there exist  $\rho^\gamma \in L^1(\Omega, X_1^*)$ ,  $\vartheta^\gamma \in \partial \mathcal{R}[J_2(x^\gamma)]$ ,  $\eta^\gamma \in \partial J_1(x_1^\gamma)$ ,  $\xi^\gamma \in N_C(x_1^\gamma)$  and  $\zeta^\gamma = (\zeta_1^\gamma, \zeta_2^\gamma) \in L^p(\Omega, X_1^*) \times L^p(\Omega, X_2^*)$  with  $\zeta^\gamma(\omega) \in \partial J_2(x_1^\gamma, x_2^\gamma(\omega); \omega)$  such that

$$\eta^\gamma + \mathbb{E}[\rho^\gamma] + \xi^\gamma = 0, \quad (39a)$$

$$\zeta_1^\gamma(\omega) \vartheta^\gamma(\omega) + e_{x_1}^*(\omega) \lambda_e^\gamma(\omega) + i_{x_1}^*(x_1^\gamma, x_2^\gamma(\omega); \omega) \lambda_i^\gamma(\omega) - \rho^\gamma(\omega) = 0, \quad (39b)$$

$$\zeta_2^\gamma(\omega) \vartheta^\gamma(\omega) + e_{x_2}^*(\omega) \lambda_e^\gamma(\omega) + i_{x_2}^*(x_1^\gamma, x_2^\gamma(\omega); \omega) \lambda_i^\gamma(\omega) = 0, \quad (39c)$$

$$x_1^\gamma \in C, \quad e(x_1^\gamma, x_2^\gamma(\omega); \omega) = 0, \quad (39d)$$

$$\lambda_i^\gamma(\omega) = \gamma(i(x_1^\gamma, x_2^\gamma(\omega); \omega) + \pi_{K_H}(-i(x_1^\gamma, x_2^\gamma(\omega); \omega))) \quad \text{in } H, \quad (39e)$$

with pointwise conditions holding for almost all  $\omega \in \Omega$ .

*Proof.* By Assumption 4.6(i), we have that the set  $C$  is contained in the induced feasible set for  $x_1$  given by  $\tilde{C} = \{x_1 \in X_1 : \exists x_2 \in L^\infty(\Omega, X_2) \text{ s.t. } e(x_1, x_2(\omega); \omega) = 0 \text{ a.s.}\}$ . Hence Assumption 3.6 is satisfied for Problem  $(P^\gamma)$  and Lemma 3.12 can be applied to the Lagrangian  $L^\gamma$ , giving the existence of a saddle point  $(x^\gamma, \lambda^\gamma)$ . Now, Theorem 3.13 can be applied.

Condition (i) of Theorem 3.13 clearly translates to (39a). As discussed in Subsection 4.1,  $\beta^\gamma$  is Fréchet differentiable on  $H$  with  $\nabla \beta^\gamma(k) = \gamma(k - \pi_{K_H}(k)) \in H \hookrightarrow R^*$ . By the chain rule,

$$\partial_{x_i} \beta^\gamma(-i(x_1, x_2; \omega)) = -i_{x_i}^*(x_1, x_2; \omega) \nabla \beta^\gamma(-i(x_1, x_2; \omega)).$$



Defining  $\lambda_i^\gamma(\omega) = -\nabla\beta^\gamma(-i(x_1, x_2; \omega))$ , we get the expression (39e). Condition (ii) of Theorem 3.13 amounts to

$$f(x_1, x_2; \omega) := J_2(x_1, x_2; \omega)\vartheta^\gamma(\omega) + \beta^\gamma(-i(x_1, x_2(\omega); \omega)) \\ + \langle \lambda_e^\gamma(\omega), e(x_1, x_2; \omega) \rangle_{W^*, W} - \langle \rho^\gamma(\omega), x_1 \rangle_{X_1^*, X_1}$$

attaining its minimum at  $(x_1^\gamma, x_2^\gamma(\omega))$ , i.e.,  $0 \in \partial f(x_1^\gamma, x_2^\gamma(\omega); \omega)$ , which yields (39b)–(39c). The remaining conditions follow directly from Theorem 3.13.  $\square$

The following characterization of the multiplier  $\lambda_i^\gamma$  will be useful later.

**Lemma 4.11.** For all  $k \in K_H$ , we have

$$(\lambda_i^\gamma(\omega), k)_H \geq 0 \quad (40)$$

and

$$(\lambda_i^\gamma(\omega), i(x_1^\gamma, x_2^\gamma(\omega); \omega))_H \geq -(\lambda_i^\gamma(\omega), k)_H$$

*Proof.* Thanks to (32) and (39e), we have  $\lambda_i^\gamma(\omega) \in K_H^\oplus$  a.s. and thus the expression (40) holds. Additionally, (32) implies  $(\lambda_i^\gamma(\omega), \pi_{K_H}(-i(x_1^\gamma, x_2^\gamma(\omega); \omega)))_H = 0$ , so

$$(\lambda_i^\gamma(\omega), k)_H \geq (\lambda_i^\gamma(\omega), \pi_{K_H}(-i(x_1^\gamma, x_2^\gamma(\omega); \omega)))_H. \quad (41)$$

Adding  $(\lambda_i^\gamma(\omega), i(x_1^\gamma, x_2^\gamma(\omega); \omega))_H$  to both sides of (41) gives

$$(\lambda_i^\gamma(\omega), i(x_1^\gamma, x_2^\gamma(\omega); \omega) + k)_H \geq 0.$$

$\square$

Let  $\mathcal{X} := X \times \Lambda \times L^1(\Omega, X_1^*) \times X_1^* \times X_1^* \times (L^1(\Omega, X_1^*) \times L^1(\Omega, X_2^*))$ . In the following, we use the shorthand  $\phi^\gamma := (\phi_1^\gamma, \phi_2^\gamma) = (\zeta_1^\gamma \vartheta^\gamma, \zeta_2^\gamma \vartheta^\gamma)$ . We define the primal-dual path

$$\mathcal{C}_r := \{(x^\gamma, \lambda^\gamma, \rho^\gamma, \eta^\gamma, \xi^\gamma, \phi^\gamma) \in \mathcal{X} : \gamma \in [r, \infty)\}.$$

**Proposition 4.12.** Let Assumption 3.1, Assumption 3.5, and Assumption 4.6(i)–4.6(ii) hold. Then, for every  $r \geq 0$ , the path  $\mathcal{C}_r$  is bounded in  $\mathcal{X}$ .

*Proof.* *Bounding of  $x^\gamma, \eta^\gamma, \phi^\gamma$ .* Making the same arguments as in Proposition 4.8,  $x^\gamma$  is bounded independently of  $\gamma$ , from which boundedness of  $\eta^\gamma$  immediately follows. Notice that  $\vartheta^\gamma \in \partial\mathcal{R}[J_2(x^\gamma)]$ , implying boundedness of  $\vartheta^\gamma$ . Using the relations (4) and (5), since  $\mathbb{E}[J_2(x^\gamma)\vartheta^\gamma] = \mathcal{R}[J_2(x^\gamma)] + \mathcal{R}^*[\vartheta^\gamma]$ , and by Remark 3.11,  $\phi^\gamma$  (after identification) belongs to  $\partial\mathbb{E}[J_2(x^\gamma)\vartheta^\gamma]$ , boundedness of  $x^\gamma$  and  $\vartheta^\gamma$  implies boundedness of  $\phi^\gamma$ .

*Bounding of  $\lambda_i^\gamma$ .* In the following, statements written as depending on  $\omega$  are to be understood as holding almost surely. From (10), there exists  $r > 0$  such that for all  $u = (u_e, u_i) \in B_r(0) = B_{r,e}(0) \times B_{r,i}(0) \subset L^\infty(\Omega, W) \times L^\infty(\Omega, R)$ , there exist  $k \in K$  and  $x^0 = (x_1^0, x_2^0(\cdot))$  satisfying  $x_1^0 \in C$  and

$$e_{x_2}(\omega)x_2^0(\omega) + e_{x_1}(\omega)x_1^0 - u_e(\omega) = 0 \\ -i(x_1^0, x_2^0(\omega); \omega) - u_i(\omega) = k. \quad (42)$$

Now, (42) combined with Lemma 4.11 gives

$$\begin{aligned}
(\lambda_i^\gamma(\omega), u_i(\omega))_H &= (\lambda_i^\gamma(\omega), i(x_1^0, x_2^0(\omega); \omega) + u_i(\omega))_H - (\lambda_i^\gamma(\omega), i(x_1^0, x_2^0(\omega); \omega))_H \\
&= -(\lambda_i^\gamma(\omega), k)_H - (\lambda_i^\gamma(\omega), i(x_1^0, x_2^0(\omega); \omega))_H \\
&\leq (\lambda_i^\gamma(\omega), i(x_1^\gamma, x_2^\gamma(\omega); \omega))_H - (\lambda_i^\gamma(\omega), i(x_1^0, x_2^0(\omega); \omega))_H \\
&\leq (\lambda_i^\gamma(\omega), i_{x_1}(x_1^\gamma, x_2^\gamma(\omega); \omega)(x_1^\gamma - x_1^0))_H \\
&\quad + (\lambda_i^\gamma(\omega), i_{x_2}(x_1^\gamma, x_2^\gamma(\omega); \omega)(x_2^\gamma(\omega) - x_2^0(\omega)))_H,
\end{aligned}$$

where we used convexity of  $(x_1, x_2) \mapsto (\lambda_i^\gamma, i(x_1, x_2, \omega))_H$  in the second inequality. Recalling (29) and using (39b)–(39d), we obtain

$$\langle \lambda_i^\gamma(\omega), u_i(\omega) \rangle_{R^*, R} \leq \langle \rho^\gamma(\omega) - \phi_1^\gamma(\omega), x_1^\gamma - x_1^0 \rangle_{X_1^*, X_1} - \langle \phi_2^\gamma(\omega), x_2^\gamma(\omega) - x_2^0(\omega) \rangle_{X_2^*, X_2}.$$

Thus

$$\begin{aligned}
\langle \lambda_i^\gamma, u_i \rangle_{L^1(\Omega, R^*), L^\infty(\Omega, R)} &\leq \langle \mathbb{E}[\rho^\gamma - \phi_1^\gamma], x_1^\gamma - x_1^0 \rangle_{X_1^*, X_1} - \mathbb{E}[\langle \phi_2^\gamma, x_2^\gamma - x_2^0 \rangle_{X_2^*, X_2}] \\
&\leq \langle -\xi^\gamma - \eta^\gamma - \mathbb{E}[\phi_1^\gamma], x_1^\gamma - x_1^0 \rangle_{X_1^*, X_1} - \mathbb{E}[\langle \phi_2^\gamma, x_2^\gamma - x_2^0 \rangle_{X_2^*, X_2}] \\
&\leq \langle -\eta^\gamma - \mathbb{E}[\phi_1^\gamma], x_1^\gamma - x_1^0 \rangle_{X_1^*, X_1} - \mathbb{E}[\langle \phi_2^\gamma, x_2^\gamma - x_2^0 \rangle_{X_2^*, X_2}] \quad (43) \\
&\leq (\|\eta^\gamma\|_{X_1^*} + \|\phi_1^\gamma\|_{L^1(\Omega, X_1^*)}) \|x_1^\gamma - x_1^0\|_{X_1} \\
&\quad + \|\phi_2^\gamma\|_{L^1(\Omega, X_2^*)} \|x_2^\gamma - x_2^0\|_{L^\infty(\Omega, X_2)},
\end{aligned}$$

where in the third inequality we used  $\xi^\gamma \in N_C(x_1^\gamma)$ , and in the fourth inequality we used Jensen's and Hölder's inequalities.

In the following, we use a generic constant  $c > 0$ , which is independent of  $\omega$  and takes potentially different values at each use. Since  $x^\gamma$ ,  $\eta^\gamma$ , and  $\phi^\gamma$  are bounded independently of  $\gamma$ , (43) yields

$$\langle \lambda_i^\gamma, u_i \rangle_{L^1(\Omega, R^*), L^\infty(\Omega, R)} \leq c < \infty$$

and hence

$$\|\lambda_i^\gamma\|_{L^1(\Omega, R^*)} \leq \frac{1}{r} \sup_{y \in B_{r,i}(0)} \{ \langle \lambda_i^\gamma, y \rangle_{L^1(\Omega, R^*), L^\infty(\Omega, R)} \} \leq \frac{1}{r} c < \infty.$$

*Bounding of  $\lambda_e^\gamma$ ,  $\rho^\gamma$ ,  $\xi^\gamma$ .* Now, (39c) and Assumption 4.6(i)–4.6(ii) yield

$$\begin{aligned}
\|\lambda_e^\gamma\|_{L^1(\Omega, W^*)} &= \|-e_{x_2}^{-*}(i_{x_2}^*(x^\gamma)\lambda_i^\gamma + \phi_2^\gamma)\|_{L^1(\Omega, W^*)} \\
&\leq c\|i_{x_2}^*(x^\gamma)\lambda_i^\gamma + \phi_2^\gamma\|_{L^1(\Omega, X_2^*)} \\
&\leq c(\|\lambda_i^\gamma\|_{L^1(\Omega, R^*)} + \|\phi_2^\gamma\|_{L^1(\Omega, X_2^*)}) \leq c < \infty.
\end{aligned}$$

Additionally, (39b) gives

$$\begin{aligned}
\|\rho^\gamma\|_{L^1(\Omega, X_1^*)} &= \|\phi_1^\gamma + e_{x_1}^*\lambda_e^\gamma + i_{x_1}^*(x^\gamma)\lambda_i^\gamma\|_{L^1(\Omega, X_1^*)} \\
&\leq c(\|\phi_1^\gamma\|_{L^1(\Omega, X_1^*)} + \|\lambda_e^\gamma\|_{L^1(\Omega, W^*)} + \|\lambda_i^\gamma\|_{L^1(\Omega, R^*)}) \leq c < \infty.
\end{aligned}$$

Finally, (39a) combined with Jensen's inequality yields

$$\|\xi^\gamma\|_{X_1^*} \leq \|\eta^\gamma\|_{X_1^*} + \mathbb{E}[\|\rho^\gamma\|_{X_1^*}] \leq c < \infty.$$

□

We now show optimality of accumulation points of the primal-dual path  $\mathcal{C}_0$ . Here, a central difficulty is that the Banach–Alagolu theorem gives weak\* compactness of a closed bounded subset of  $L^1(\Omega, X^*)$  ( $X$  being a generic reflexive Banach space) as a dual of another Banach space. It is not clear what form the predual takes. Another difficulty is that weakly\* compact subsets of  $L^1(\Omega, X^*)$  are not weakly\* sequentially compact.

We will use two tools to circumvent these difficulties: first, we will identify  $L^1(\Omega, X^*)$  with a closed subspace of the bidual space  $(L^1(\Omega, X^*))^{**} = (L^\infty(\Omega, X))^*$ . This is always possible, and in our setting, this corresponds to the relationship between absolutely continuous functionals and their integrable functionals via the expression (2). The absolutely continuous functional corresponding to  $\lambda^\gamma$  is denoted by  $\lambda^{\gamma,a}$ ; the other functionals are defined similarly.

To circumvent the second difficulty, we will work with a *net* (generalized sequence) in  $\mathcal{C}_0$  instead of a sequence. Given a directed set  $I$  on which a relation  $\preceq$  is defined, which is reflexive, transitive, and directed, a net  $(x^\alpha)_{\alpha \in I}$  on  $X$  is a map  $x : I \rightarrow X$ . A net  $(y^\beta)_{\beta \in J}$  is a subnet of  $(x^\alpha)_{\alpha \in I}$  via  $\varphi : J \rightarrow I$  if for all  $\beta \in J$ ,  $y^\beta = x^{\varphi(\beta)}$  and for every  $\alpha_0 \in I$ , there exists  $\beta_0 \in J$  for which  $\beta_0 \preceq \beta$  implies  $\alpha_0 \preceq \varphi(\beta)$ . We remind that  $x^\alpha$  converges to an accumulation point  $x$  if  $(x^\alpha)_{\alpha \in I}$  lies eventually in every neighborhood of  $x$ , i.e., for any neighborhood  $V$  of  $x$ , there exists  $\alpha_0 \in I$  such that for all  $\alpha \succeq \alpha_0$ , we have  $x^\alpha \in V$ . For our purposes, the choice  $(I, \preceq) = ([0, \infty), \leq)$  will be sufficient. Notably, topological notions of closure and continuity carry over for nets. We will write  $\rightarrow$ ,  $\rightharpoonup$ ,  $\rightharpoonup^*$  to emphasize topologies when taking limits. Nets on compact sets in Hausdorff spaces have the following convenient property: there exists a subnet that converges to a point in the set; see, e.g., [7, Section 1.7] for definitions and properties.

**Theorem 4.13.** Let Assumption 3.1, Assumption 3.5, and Assumption 4.6 hold, and observe the net generated by the primal-dual path  $(x^{\gamma'}, \lambda^{\gamma'}, \rho^{\gamma'}, \eta^{\gamma'}, \xi^{\gamma'}, \phi^{\gamma'})_{\gamma' \in [0, \infty)} \subset \mathcal{C}_0$ , with  $\phi^{\gamma'} = \zeta^{\gamma'} \vartheta^{\gamma'}$ . Then there exists for  $J \subset [0, \infty)$  a subnet  $(x^\gamma, \lambda^{\gamma,a}, \rho^{\gamma,a}, \eta^\gamma, \xi^\gamma, \phi^{\gamma,a})_{\gamma \in J}$ , a sequence  $\gamma_n \subset J$  with  $\gamma_n \rightarrow \infty$ , and an accumulation point  $(\hat{x}, \lambda', \rho', \hat{\eta}, \hat{\xi}, \phi')$ ,  $\hat{\zeta} \in L^p(\Omega, X_1^*) \times L^p(\Omega, X_2^*)$ , and  $\hat{\vartheta} \in L^{p'}(\Omega)$  such that

- 1  $x^{\gamma_n} \rightarrow \hat{x}$  in  $X$ ,
- 2  $\lambda_e^{\gamma_n,a} \rightharpoonup^* \lambda'_e$  in  $(L^\infty(\Omega, W))^*$ ,
- 3  $\lambda_i^{\gamma_n,a} \rightharpoonup^* \lambda'_i$  in  $(L^\infty(\Omega, R))^*$ ,
- 4  $\rho^{\gamma_n,a} \rightharpoonup^* \rho'$  in  $(L^\infty(\Omega, X_1))^*$ ,
- 5  $\eta^{\gamma_n} \rightarrow \hat{\eta}$  in  $X_1^*$ ,
- 6  $\xi^{\gamma_n} \rightarrow \hat{\xi}$  in  $X_1^*$ ,
- 7  $\phi^{\gamma_n,a} \rightharpoonup^* \phi'$  in  $(L^\infty(\Omega, X_1))^* \times (L^\infty(\Omega, X_2))^*$  with  $\vartheta^{\gamma_n} \rightharpoonup^* \hat{\vartheta}$  in  $L^{p'}(\Omega)$  and  $\zeta^{\gamma_n}(\omega) \rightarrow \hat{\zeta}(\omega)$  in  $X_1^* \times X_2^*$  a.s.

We have  $\lambda'_e = \hat{\lambda}_e^a + \hat{\lambda}_e^\circ$ ,  $\lambda'_i = \hat{\lambda}_i^a + \hat{\lambda}_i^\circ$ ,  $\rho' = \hat{\rho}^a + \hat{\rho}^\circ$ , and  $\phi' = \hat{\phi}^a + \hat{\phi}^\circ$  for the absolutely continuous functionals  $\hat{\lambda}_e^a \in (L^\infty(\Omega, W))^*$ ,  $\hat{\lambda}_i^a \in (L^\infty(\Omega, R))^*$ ,  $\hat{\rho}^a \in (L^\infty(\Omega, X_1))^*$ , and  $\hat{\phi}^a \in (L^\infty(\Omega, X_1))^* \times (L^\infty(\Omega, X_2))^*$ , corresponding to  $\hat{\lambda}_e \in L^1(\Omega, W^*)$ ,  $\hat{\lambda}_i \in L^1(\Omega, R^*)$ ,  $\hat{\rho} \in L^1(\Omega, X_1^*)$ , and  $\hat{\phi} \in (L^1(\Omega, X_1^*) \times L^1(\Omega, X_2^*))$ , respectively. Finally, the accumulation point  $(\hat{x}, \hat{\lambda}, \hat{\lambda}^\circ, \hat{\rho}, \hat{\rho}^\circ, \hat{\eta}, \hat{\xi}, \hat{\phi}, \hat{\phi}^\circ)$  satisfies the following conditions:  $\hat{\eta} \in \partial J_1(\hat{x}_1)$ ,  $\hat{\xi} \in N_C(\hat{x}_1)$ ,  $\hat{\vartheta} \in$

$\partial\mathcal{R}[J_2(\hat{x})]$ ,  $\hat{\zeta}(\omega) \in \partial J_2(\hat{x}_1, \hat{x}_2(\omega); \omega)$  a.s., and:

$$\hat{\eta} + \mathbb{E}[\hat{\rho}] + \hat{\phi}_1^\circ + e_{x_1}^* \hat{\lambda}_e^\circ + i_{x_1}^*(\hat{x}) \hat{\lambda}_i^\circ + \hat{\xi} = 0, \quad (44a)$$

$$\hat{\phi}_2^\circ + e_{x_2}^* \hat{\lambda}_e^\circ + i_{x_2}^*(\hat{x}) \hat{\lambda}_i^\circ = 0, \quad (44b)$$

$$\hat{\phi}_1(\omega) + e_{x_1}^*(\omega) \hat{\lambda}_e(\omega) + i_{x_1}^*(\hat{x}_1, \hat{x}_2(\omega); \omega) \hat{\lambda}_i(\omega) - \hat{\rho}(\omega) = 0, \quad (44c)$$

$$\hat{\phi}_2(\omega) + e_{x_2}^*(\omega) \hat{\lambda}_e(\omega) + i_{x_2}^*(\hat{x}_1, \hat{x}_2(\omega); \omega) \hat{\lambda}_i(\omega) = 0, \quad (44d)$$

$$\hat{x}_1 \in C, \quad e(\hat{x}_1, \hat{x}_2(\omega); \omega) = 0, \quad (44e)$$

$$i(\hat{x}_1, \hat{x}_2(\omega); \omega) \leq_K 0, \quad \hat{\lambda}_i(\omega) \in K^\oplus, \quad \langle \hat{\lambda}_i(\omega), i(\hat{x}_1, \hat{x}_2(\omega); \omega) \rangle_{R^*, R} = 0, \quad (44f)$$

$$\hat{\lambda}_i^\circ \in \mathcal{K}^\oplus, \quad \langle \hat{\lambda}_i^\circ, i(\hat{x}) \rangle_{(L^\infty(\Omega, R))^*, L^\infty(\Omega, R)} = 0, \quad (44g)$$

with pointwise conditions holding for almost all  $\omega \in \Omega$ .

*Proof. Characterization of accumulation points and topologies.* By Proposition 4.12, for every  $r \geq 0$ , the path  $\mathcal{C}_r$  is bounded in  $\mathcal{X}$ , in particular for  $r = 0$ . Thus  $\mathcal{C}_0$  is weakly\* compact and hence there exists a subnet  $(x^\gamma, \lambda^\gamma, \rho^\gamma, \eta^\gamma, \xi^\gamma, \phi^\gamma)_{\gamma \in J} \subset \mathcal{C}_0$  such that (after identifying absolutely continuous terms), we have  $(x^\gamma, \lambda^{\gamma, a}, \rho^{\gamma, a}, \eta^\gamma, \xi^\gamma, \phi^{\gamma, a}) \rightharpoonup^* (\hat{x}, \lambda', \rho', \hat{\eta}, \hat{\xi}, \phi')$ . Note that the decomposition of accumulation points, i.e.,  $\lambda'_e = \hat{\lambda}_e^a + \hat{\lambda}_e^\circ$ ,  $\lambda'_i = \hat{\lambda}_i^a + \hat{\lambda}_i^\circ$ ,  $\rho' = \hat{\rho}^a + \hat{\rho}^\circ$ , and  $\phi' = \hat{\phi}^a + \hat{\phi}^\circ$  follows from Theorem 2.4.

One can prove  $j(x^\gamma) \rightarrow j(\hat{x})$  for the subnet  $(x^\gamma)_{\gamma \in J}$  exactly as in Proposition 4.8; see in particular (37). Together with Assumption 4.6(iii), we obtain  $\|x_1^\gamma\|_{X_1} \rightarrow \|\hat{x}_1\|_{X_1}$  as  $n \rightarrow \infty$  and thus  $x_1^\gamma \rightarrow \hat{x}_1$  in  $X_1$ . From the formula (33) we have

$$\operatorname{ess\,sup}_{\omega \in \Omega} \|x_2^\gamma(\omega) - \hat{x}_2(\omega)\|_{X_2} = \operatorname{ess\,sup}_{\omega \in \Omega} \|e_{x_2}^{-1}(\omega) e_{x_1}(\omega) (x_1^\gamma - \hat{x}_1)\|_{X_2},$$

and hence  $x_2^\gamma \rightarrow \hat{x}_2$  in  $L^\infty(\Omega, X_2)$ . We obtain  $x^\gamma \rightarrow \hat{x}$  in  $X$ .

*Verification of (44c).* From (39b), we have for all  $\mu \in L^\infty(\Omega, X_1)$  and all  $\gamma$

$$0 = \langle \phi_1^\gamma + e_{x_1}^*(\cdot) \lambda_e^\gamma(\cdot) + i_{x_1}^*(x_1^\gamma, x_2^\gamma(\cdot); \cdot) \lambda_i^\gamma(\cdot) - \rho^\gamma, \mu \rangle_{L^1(\Omega, X_1^*), L^\infty(\Omega, X_1)}.$$

Notice that

- $\langle \phi_1^\gamma, \mu \rangle_{L^1(\Omega, X_1^*), L^\infty(\Omega, X_1)} = \langle \phi_1^{\gamma, a}, \mu \rangle_{(L^\infty(\Omega, X_1))^*, L^\infty(\Omega, X_1)} \rightarrow \langle \phi_1', \mu \rangle_{(L^\infty(\Omega, X_1))^*, L^\infty(\Omega, X_1)}$  by weak\* convergence of  $\phi_1^\gamma$  in  $(L^\infty(\Omega, X_1))^*$ .

- Using the identity for absolutely continuous functionals (2) combined with Lemma 2.6 we have

$$\begin{aligned} \langle e_{x_1}^*(\cdot) \lambda_e^\gamma(\cdot), \mu \rangle_{L^1(\Omega, X_1^*), L^\infty(\Omega, X_1)} &= \langle \lambda_e^\gamma, e_{x_1}(\cdot) \mu(\cdot) \rangle_{L^1(\Omega, W), L^\infty(\Omega, W)} \\ &= \langle \lambda_e^{\gamma, a}, e_{x_1}(\cdot) \mu(\cdot) \rangle_{(L^\infty(\Omega, W))^*, L^\infty(\Omega, W)} \\ &\rightarrow \langle \hat{\lambda}_e, e_{x_1}(\cdot) \mu(\cdot) \rangle_{(L^\infty(\Omega, W))^*, L^\infty(\Omega, W)} = \langle e_{x_1}^* \hat{\lambda}_e, \mu \rangle_{(L^\infty(\Omega, X_1))^*, L^\infty(\Omega, X_1)}, \end{aligned}$$

by weak\* convergence of  $\lambda_e^{\gamma, a}$  in  $(L^\infty(\Omega, W))^*$ .

- From strong convergence of  $(x^\gamma)_{\gamma \in J}$  and weak\* convergence of  $\lambda_i^{\gamma, a}$  in  $(L^\infty(\Omega, R))^*$  we have

$$\begin{aligned} \langle i_{x_1}^*(x_1^\gamma, x_2^\gamma(\cdot); \cdot) \lambda_i^\gamma(\cdot), \mu \rangle_{L^1(\Omega, X_1^*), L^\infty(\Omega, X_1)} &= \langle \lambda_i^\gamma(\cdot), i_{x_1}(x_1^\gamma, x_2^\gamma(\cdot); \cdot) \mu(\cdot) \rangle_{L^1(\Omega, R^*), L^\infty(\Omega, R)} \\ &= \langle \lambda_i^{\gamma, a}, i_{x_1}(x_1^\gamma, x_2^\gamma(\cdot); \cdot) \mu(\cdot) \rangle_{(L^\infty(\Omega, R))^*, L^\infty(\Omega, R)} \\ &\rightarrow \langle \hat{\lambda}_i, i_{x_1}(\hat{x}_1, \hat{x}_2(\cdot); \cdot) \mu(\cdot) \rangle_{(L^\infty(\Omega, R))^*, L^\infty(\Omega, R)} = \langle i_{x_1}^*(\hat{x}_1, \hat{x}_2(\cdot); \cdot) \hat{\lambda}_i, \mu \rangle_{(L^\infty(\Omega, X_1))^*, L^\infty(\Omega, X_1)}, \end{aligned}$$

where we used the continuity of the partial derivative afforded by Assumption 4.6(ii).

- $\langle \rho^\gamma, \mu \rangle_{L^1(\Omega, X_1^*), L^\infty(\Omega, X_1)} = \langle \rho^{\gamma, a}, \mu \rangle_{(L^\infty(\Omega, X_1))^*, L^\infty(\Omega, X_1)} \rightarrow \langle \rho^{\gamma, a}, \mu \rangle_{(L^\infty(\Omega, X_1))^*, L^\infty(\Omega, X_1)}$   
by weak\* convergence of  $\rho^{\gamma, a}$  in  $(L^\infty(\Omega, X_1))^*$ .

Thus

$$0 = \langle \phi'_1 + e_{x_1}^* \lambda'_e + i_{x_1}^*(\hat{x}) \lambda'_i - \rho', \mu \rangle_{(L^\infty(\Omega, X_1))^*, L^\infty(\Omega, X_1)}.$$

The sets  $\mathcal{A}_1$  and  $\mathcal{S}_1$  denote the sets of absolutely continuous functionals and singular functionals, respectively, defined on  $L^\infty(\Omega, X_1)$ . By Lemma 2.6, it follows that

$$\underbrace{(\hat{\phi}_1^\circ + e_{x_1}^* \hat{\lambda}_e^\circ + i_{x_1}^*(\hat{x}) \hat{\lambda}_i^\circ - \hat{\rho}^\circ)}_{\in \mathcal{S}_1} + \underbrace{(\hat{\phi}_1^a + e_{x_1}^* \hat{\lambda}_e^a + i_{x_1}^*(\hat{x}) \hat{\lambda}_i^a - \hat{\rho}^a)}_{\in \mathcal{A}_1 \cong L^1(\Omega, X_1^*)} = 0.$$

The spaces  $\mathcal{A}_1$  and  $\mathcal{S}_1$  are complementary spaces, so  $\mathcal{A}_1 \cap \mathcal{S}_1 = \{0\}$ . This implies

$$\hat{\phi}_1^\circ + e_{x_1}^* \hat{\lambda}_e^\circ + i_{x_1}^*(\hat{x}) \hat{\lambda}_i^\circ - \hat{\rho}^\circ = 0, \quad (45)$$

$$\hat{\phi}_1^a + e_{x_1}^* \hat{\lambda}_e^a + i_{x_1}^*(\hat{x}) \hat{\lambda}_i^a - \hat{\rho}^a = 0. \quad (46)$$

The expression (46) immediately implies that (44c) is satisfied.

*Verification of (44b) and (44d).* Analogously to the prior step, we have from (39c) for all  $\nu \in L^\infty(\Omega, X_2)$  and all  $\gamma$

$$0 = \langle \phi_2^\gamma + e_{x_2}^*(\cdot) \lambda_e^\gamma(\cdot) + i_{x_2}^*(x_1^\gamma, x_2^\gamma(\cdot); \cdot) \lambda_i^\gamma(\cdot), \nu \rangle_{L^1(\Omega, X_2^*), L^\infty(\Omega, X_2)}.$$

Hence

$$0 = \langle \phi'_2 + e_{x_2}^* \lambda'_e + i_{x_2}^*(\hat{x}) \lambda'_i, \nu \rangle_{(L^\infty(\Omega, X_2))^*, L^\infty(\Omega, X_2)}.$$

Grouping by singular and absolutely continuous terms, we obtain (44b) and (44d).

*Verification of (44a) and (44e).* The expression (44e) follows from Corollary 2.7 and the fact that  $C$  is closed. From (39a), we have for all  $\mu \in X_1$

$$\begin{aligned} 0 &= \langle \mathbb{E}[\rho^\gamma] + \eta^\gamma + \xi^\gamma, \mu \rangle_{X_1^*, X_1} \\ &= \langle \rho^{\gamma, a}, \mu \rangle_{(L^\infty(\Omega, X_1))^*, L^\infty(\Omega, X_1)} + \langle \eta^\gamma + \xi^\gamma, \mu \rangle_{X_1^*, X_1}. \end{aligned}$$

After plugging in (45), the accumulation point therefore satisfies

$$\begin{aligned} 0 &= \langle \hat{\rho}^a + \hat{\rho}^\circ, \mu \rangle_{(L^\infty(\Omega, X_1))^*, L^\infty(\Omega, X_1)} + \langle \hat{\eta} + \hat{\xi}, \mu \rangle_{X_1^*, X_1} \\ &= \langle \mathbb{E}[\hat{\rho}] + \hat{\phi}_1^\circ + e_{x_1}^* \hat{\lambda}_e^\circ + i_{x_1}^*(\hat{x}) \hat{\lambda}_i^\circ + \hat{\eta} + \hat{\xi}, \mu \rangle_{X_1^*, X_1} \end{aligned}$$

for all  $\mu \in X_1$ , which is equivalent to (44a).

*Verification of (44f) and (44g).* We proved  $i(\hat{x}_1, \hat{x}_2(\omega); \omega) \leq_K 0$  in Proposition 4.8. For all  $\gamma, \lambda_i^\gamma(\omega) \in K_H^\oplus$ . It follows for all  $r \in L^\infty(\Omega, R)$  satisfying  $r(\omega) \in K_H \cap R = K$  that

$$0 \leq \mathbb{E}[(\lambda_i^\gamma, r)_H] = \mathbb{E}[(\lambda_i^\gamma, r)_{R^*, R}] = \langle \lambda_i^{\gamma, a}, r \rangle_{(L^\infty(\Omega, R))^*, L^\infty(\Omega, R)}.$$

Therefore

$$0 \leq \langle \hat{\lambda}_i^a + \hat{\lambda}_i^\circ, r \rangle_{(L^\infty(\Omega, R))^*, L^\infty(\Omega, R)}.$$

Since  $\hat{\lambda}_i^a$  and  $\hat{\lambda}_i^\circ$  belong to complementary subspaces, we conclude

$$\hat{\lambda}_i(\omega) \in K^\oplus \text{ a.s. and } \hat{\lambda}_i^\circ \in K^\oplus. \quad (47)$$

By weak\* convergence of  $\lambda_i^\gamma$ , strong convergence of  $x^\gamma$ , and continuity of  $x \mapsto i(x_1, x_2(\cdot); \cdot)$ , we have

$$\begin{aligned} \mathbb{E}[(\lambda_i^\gamma, -i(x_1^\gamma, x_2^\gamma(\cdot); \cdot))_H] &= \langle \lambda_i^{\gamma, a}, -i(x_1^\gamma, x_2^\gamma(\cdot); \cdot) \rangle_{(L^\infty(\Omega, R))^*, L^\infty(\Omega, R)} \\ &\rightarrow \langle \lambda_i', -i(\hat{x}_1, \hat{x}_2(\cdot); \cdot) \rangle_{(L^\infty(\Omega, R))^*, L^\infty(\Omega, R)} \geq 0. \end{aligned} \quad (48)$$

Notice that for all  $\gamma$ ,

$$\begin{aligned} \mathbb{E}[(\lambda_i^\gamma, -i(x^\gamma))_H] &= \gamma \mathbb{E}[(i(x^\gamma) + \pi_{K_H}(-i(x^\gamma))), -i(x^\gamma)]_H \\ &= -\gamma \mathbb{E}[\|i(x^\gamma)\|_H^2] + \gamma \mathbb{E}[(\pi_{K_H}(-i(x^\gamma)) - \pi_{K_H}(0), -i(x^\gamma))_H] \\ &\leq -\gamma \mathbb{E}[\|i(x^\gamma)\|_H^2] + \gamma \mathbb{E}[\|\pi_{K_H}(-i(x^\gamma)) - \pi_{K_H}(0)\|_H \|i(x^\gamma)\|_H] \\ &\leq -\gamma \mathbb{E}[\|i(x^\gamma)\|_H^2] + \gamma \mathbb{E}[\|i(x^\gamma)\|_H^2] = 0, \end{aligned}$$

where we used monotonicity of the expectation operator, nonexpansivity of the projection operator, as well as  $0 \in K_H$ . This combined with (48) yields

$$\langle \hat{\lambda}_i(\omega), i(\hat{x}_1, \hat{x}_2(\omega); \omega) \rangle_{R^*, R} = 0, \quad \text{and} \quad \langle \hat{\lambda}_i^\circ, i(\hat{x}) \rangle_{(L^\infty(\Omega, R))^*, L^\infty(\Omega, R)} = 0.$$

*Sequential statements.* We note that  $X_1$  and  $L^\infty(\Omega, X_2)$  are metrizable spaces. Note that  $X_1$  is reflexive. From the subnet  $(x^\gamma, \eta^\gamma, \xi^\gamma)_{\gamma \in J}$ , it is therefore possible to extract a subsequence  $\gamma_n \rightarrow \infty$  such that  $x^{\gamma_n} \rightarrow \hat{x}$ ,  $\eta^{\gamma_n} \rightarrow \hat{\eta}$ , and  $\xi^{\gamma_n} \rightarrow \hat{\xi}$ .

*Convergence of elements from subdifferentials.* Strong convergence of  $x^{\gamma_n}$  and weak convergence of  $\xi^{\gamma_n} \in N_C(x_1^{\gamma_n})$  imply  $\hat{\xi} \in N_C(\hat{x}_1)$ . By the same reasoning,  $\hat{\eta} \in \partial J_1(\hat{x}_1)$ . We now reveal the structure of  $\phi^{\gamma_n} = \vartheta^{\gamma_n} \zeta^{\gamma_n}$ , where  $\vartheta^{\gamma_n} \in \partial \mathcal{R}[J_2(x^{\gamma_n})]$  and  $\zeta^{\gamma_n}(\omega) \in \partial J_2(x_1^{\gamma_n}, x_2^{\gamma_n}(\omega); \omega)$  a.s. Clearly,  $\vartheta^{\gamma_n}$  is bounded in  $L^{p'}(\Omega)$  due to the boundedness of  $x^{\gamma_n}$ , hence on a subsequence (we use the same labeling) there exists a  $\hat{\vartheta} \in L^{p'}(\Omega)$  such that  $\vartheta^{\gamma_n} \rightharpoonup^* \hat{\vartheta}$ . Therefore, strong convergence of  $x^{\gamma_n}$  to  $\hat{x}$  implies  $\hat{\vartheta} \in \partial \mathcal{R}[J_2(\hat{x})]$ . Moreover, for almost every  $\omega$ , the sequence  $\zeta^{\gamma_n}(\omega)$  is bounded in  $X_1^* \times X_2^*$  due to the boundedness of  $x^{\gamma_n}$ , hence there exists a subsequence  $\{n_k\}$  and a limit point  $\hat{\zeta}(\omega)$  such that  $\zeta^{\gamma_{n_k}}(\omega) \rightharpoonup \hat{\zeta}(\omega)$ . From the convergence  $x^{\gamma_n} \rightarrow \hat{x}$ , we have on a further subsequence (with the same labeling) the pointwise convergence  $x^{\gamma_{n_k}}(\omega) \rightarrow \hat{x}(\omega)$  a.s. and hence  $\hat{\zeta}(\omega) \in \partial J_2(\hat{x}_1, \hat{x}_2(\omega); \omega)$  a.s. The fact that  $\hat{\zeta} \in L^p(\Omega, X_1^*) \times L^p(\Omega, X_2^*)$  follows by Remark 3.11.  $\square$

The system (44) in Theorem 4.13 is not necessary and sufficient for optimality since it is not generally guaranteed that the sequence given by  $\phi^{\gamma_n} = \vartheta^{\gamma_n} \zeta^{\gamma_n}$  has a limit of the form  $\hat{\vartheta} \hat{\zeta}$  as required by (38c)–(38d). Moreover, the system (44) contains the singular terms  $\hat{\phi}^\circ$ . However, there are two relevant cases in which it is possible to show the stronger result.

**Corollary 4.14.** With the same assumptions as in Theorem 4.13, assume additionally that  $\mathcal{R} = \mathbb{E}$ . Then  $\hat{\phi} = \hat{\zeta}$ ,  $\hat{\phi}^\circ \equiv 0$ , and the system (44) is necessary and sufficient for optimality.

*Proof.* In the case where  $\mathcal{R} = \mathbb{E}$ , we have  $\vartheta^\gamma = \hat{\vartheta} \equiv 1$  for all  $\gamma \geq 0$ . Hence from  $\hat{\zeta}(\omega) \in \partial J_2(\hat{x}_1, \hat{x}_2(\omega); \omega)$  a.s. and  $\phi^{\gamma_n} = \zeta^{\gamma_n}$  (after identification) belongs to  $\partial \mathbb{E}[J_2(x^{\gamma_n})]$ , which using strong convergence of  $x^{\gamma_n}$  implies that  $\hat{\phi} \in \partial \mathbb{E}[J_2(\hat{x})]$ . From Lemma 2.5, we obtain that  $\hat{\phi}^\circ \equiv 0$ . By Lemma 4.9, the conditions posed by (44) are now necessary and sufficient.  $\square$

**Corollary 4.15.** With the same assumptions as in Theorem 4.13, assume additionally that  $J_2(\cdot, \cdot; \omega)$  is continuously Fréchet differentiable on  $X_1 \times X_2$  with derivative  $DJ_2(\cdot, \cdot; \omega)$  for almost every  $\omega$ ; for every  $r > 0$  there exists  $a'_r \in L^p(\Omega)$  such that  $DJ_2(x_1, x_2; \omega) \leq a'_r(\omega)$  a.s. for  $\|x_1\|_{X_1} + \|x_2\|_{X_2} \leq r$ . Then  $\hat{\phi} = \hat{\vartheta} \hat{\zeta}$ ,  $\hat{\phi}^\circ \equiv 0$ , and the system (44) is necessary and sufficient for optimality.

*Proof.* Fréchet differentiability of  $J_2$  implies the subdifferential is a singleton, i.e.,  $\partial J_2(x_1, x_2; \omega) = \{DJ_2(x_1, x_2; \omega)\}$ . From strong convergence  $x^{\gamma_n} \rightarrow \hat{x}$  in  $X$  one has (potentially on a subsequence, where we use the same labeling)  $(x_1^{\gamma_n}, x_2^{\gamma_n}(\omega)) \rightarrow (\hat{x}_1, \hat{x}_2(\omega))$  a.s. In particular it follows by continuity of the derivative that

$$\zeta^{\gamma_n}(\omega) = DJ_2(x_1^{\gamma_n}, x_2^{\gamma_n}(\omega); \omega) \rightarrow DJ_2(\hat{x}_1, \hat{x}_2(\omega); \omega) = \hat{\zeta}(\omega) \quad \text{in } X_1^* \times X_2^* \text{ a.s.}$$

Together with the growth condition and Lebesgue's Dominated Convergence theorem, we have strong convergence  $\zeta^{\gamma_n} \rightarrow \hat{\zeta}$  in  $L^p(\Omega, X_1^*) \times L^p(\Omega, X_2^*)$ . We now prove that  $\phi^{\gamma_n} \rightarrow \hat{\phi} = \hat{\vartheta} \hat{\zeta} \in \mathcal{Y} := L^1(\Omega, X_1^*) \times L^1(\Omega, X_2^*)$ . Naturally, we have

$$\langle v, \phi^{\gamma_n} \rangle_{\mathcal{Y}^*, \mathcal{Y}} = \langle v, \vartheta^{\gamma_n}(\zeta^{\gamma_n} - \hat{\zeta}) \rangle_{\mathcal{Y}^*, \mathcal{Y}} + \langle v, \vartheta^{\gamma_n} \hat{\zeta} \rangle_{\mathcal{Y}^*, \mathcal{Y}} \quad \forall v \in \mathcal{Y}^*. \quad (49)$$

Notice that

$$\langle v, \vartheta^{\gamma_n}(\zeta^{\gamma_n} - \hat{\zeta}) \rangle_{\mathcal{Y}^*, \mathcal{Y}} \leq \|v\|_{\mathcal{Y}^*} (\|\vartheta^{\gamma_n}(\zeta_1^{\gamma_n} - \hat{\zeta}_1)\|_{L^1(\Omega, X_1^*)} + \|\vartheta^{\gamma_n}(\zeta_2^{\gamma_n} - \hat{\zeta}_2)\|_{L^1(\Omega, X_2^*)})$$

and for  $i = 1, 2$ ,

$$\begin{aligned} \|\vartheta^{\gamma_n}(\zeta_i^{\gamma_n} - \hat{\zeta}_i)\|_{L^1(\Omega, X_i^*)} &= \int_{\Omega} \|\vartheta^{\gamma_n}(\omega)(\zeta_i^{\gamma_n}(\omega) - \hat{\zeta}_i(\omega))\|_{X_i^*} \mathbf{d}\mathbb{P}(\omega) \\ &= \int_{\Omega} |\vartheta^{\gamma_n}(\omega)| \|(\zeta_i^{\gamma_n}(\omega) - \hat{\zeta}_i(\omega))\|_{X_i^*} \mathbf{d}\mathbb{P}(\omega) \\ &\leq \|\vartheta^{\gamma_n}\|_{L^{p'}(\Omega)} \|\zeta_i^{\gamma_n} - \hat{\zeta}_i\|_{L^p(\Omega, X_i^*)}, \end{aligned}$$

where we used Hölder's inequality in the last step. Combining this with strong convergence  $\zeta^{\gamma_n} \rightarrow \hat{\zeta}$  and boundedness of  $\vartheta^{\gamma_n}$ , the first term on the right hand side of (49) disappears in the limit. For the second term on the right hand side of (49), we have

$$\begin{aligned} \langle v, \vartheta^{\gamma_n} \hat{\zeta} \rangle_{\mathcal{Y}^*, \mathcal{Y}} &= \sum_{i=1}^2 \int_{\Omega} \langle v_i(\omega), \vartheta^{\gamma_n}(\omega) \hat{\zeta}(\omega) \rangle_{X_i, X_i^*} \mathbf{d}\mathbb{P}(\omega) \\ &= \sum_{i=1}^2 \int_{\Omega} \vartheta^{\gamma_n}(\omega) \langle v_i(\omega), \hat{\zeta}(\omega) \rangle_{X_i, X_i^*} \mathbf{d}\mathbb{P}(\omega). \end{aligned}$$

Notice that  $\langle v_i(\cdot), \hat{\zeta}(\cdot) \rangle_{X_i, X_i^*} \in L^p(\Omega)$  by the regularity of  $\hat{\zeta}$ . Hence  $\vartheta^{\gamma_n} \rightarrow^* \hat{\vartheta}$  gives  $\langle v, \vartheta^{\gamma_n} \hat{\zeta} \rangle_{\mathcal{Y}^*, \mathcal{Y}} \rightarrow \langle v, \hat{\vartheta} \hat{\zeta} \rangle_{\mathcal{Y}^*, \mathcal{Y}}$  for all  $v \in \mathcal{Y}^*$ . Returning to (49), we obtain  $\langle v, \phi^{\gamma_n} \rangle_{\mathcal{Y}^*, \mathcal{Y}} \rightarrow \langle v, \hat{\phi} \rangle_{\mathcal{Y}^*, \mathcal{Y}}$  for all  $v \in \mathcal{Y}^*$ . Since  $\hat{\phi} := \hat{\vartheta} \hat{\zeta} \in \mathcal{Y}$ , the singular term  $\hat{\phi}^\circ$  vanishes, and the system (44) is necessary and sufficient for optimality.  $\square$

To close this section, we discuss the situation where the adjoint variable  $\lambda_e^\gamma$  is more regular than our strict framework would suggest. The choice of  $W$  and  $R$  clearly need to be compatible with the interior point condition (10) for the original problem. However, for Problem  $(P^\gamma)$ , the condition (10) is trivially fulfilled as long as  $e_{x_2}$  is surjective. Without the restrictions on  $W$ , it is possible to gain additional regularity in the adjoint variable  $\lambda_e^\gamma$  if we view the state  $x_2^\gamma$  in a less regular space, a fact that can be exploited in computations. In the following lemma, we make this idea more precise.

**Lemma 4.16.** Under the same conditions as Lemma 4.10, suppose further that  $X_2$  and  $W$  are continuously embedded in  $\tilde{X}_2$  and  $\tilde{W}$ , respectively. Suppose  $e(\cdot, \cdot; \omega) \in \mathcal{L}(X_1 \times \tilde{X}_2, \tilde{W})$  and  $e_{x_2}(\omega) \in \mathcal{L}(\tilde{X}_2, \tilde{W})$  is a linear isomorphism for almost every  $\omega \in \Omega$  with  $e_{x_2}^{-1} \in L^\infty(\Omega, \mathcal{L}(\tilde{W}, \tilde{X}_2))$ . Additionally, suppose  $i(\cdot, \cdot; \omega)$  is continuously Fréchet differentiable on  $X_1 \times \tilde{X}_2$  for almost every  $\omega$  with partial derivative satisfying  $i_{x_2}(x_1, x_2; \cdot) \in L^\infty(\Omega, \mathcal{L}(\tilde{X}_2, R))$ . Then  $\lambda_e^\gamma$  belongs to  $L^1(\Omega, \tilde{W}^*)$ .

*Proof.* From (39c) we immediately obtain

$$\lambda_e^\gamma(\omega) = -e_{x_2}^{-*}(\omega)(\zeta_2^\gamma(\omega)\vartheta^\gamma(\omega) + i_{x_2}^*(x_1^\gamma, x_2^\gamma(\omega); \omega)\lambda_i^\gamma(\omega)) \in \tilde{W}^*.$$

The fact that  $\lambda_e^\gamma \in L^1(\Omega, \tilde{W}^*)$  follows by the assumptions.  $\square$

## 5 Conclusion and Discussion

In this paper, we presented optimality conditions for a large class of potentially risk-averse convex stochastic optimization problems. The applications in mind involve optimal control problems, where the control belongs to a Banach space and is coupled to a state, which is a vector-valued random variable. [19] already presented optimality conditions under the assumption of relatively complete recourse, which – while not satisfiable for nontrivial problems with additional state constraints – is satisfied for a sequence of regularized problems. In this paper, we strengthened the [19] results by presenting optimality conditions without this assumption, resulting in optimality conditions having singular Lagrange multipliers. Additionally, we formulated optimality conditions general enough to handle problems having risk measures in the objective.

As the second main contribution, we showed that regularized problems using a Moreau–Yosida regularization for the conical constraints are consistent with the original problem as the penalization parameter is taken to infinity. While [19] already showed consistency with respect to the primal variables for a specific example, we generalize these results further and show consistency with respect to dual variables.

The results presented here provide the framework for constructing algorithms via a penalization approach. For algorithms, it is preferred to work with sequential statements as opposed to nets. This is possible on sets of vanishingly small measure using the following result.

**Lemma 5.1** ([6]). Suppose  $X$  is a reflexive Banach space and  $\{f_n\} \subset L^1(\Omega, X)$  is a bounded sequence. Then there exists  $\bar{f} \in L^1(\Omega, X)$  and a subsequence  $\{f_{n_j}\}$  and nonincreasing sequence of sets  $E_k \in \mathcal{F}$  with  $\lim_{k \rightarrow \infty} \mathbb{P}(E_k) = 0$  such that

$$f_{n_j} \rightharpoonup \bar{f} \quad \text{in } L^1(\Omega \setminus E_k, X)$$

for every fixed  $k$ .

A drawback of Lemma 5.1 is that one does not know the sets  $E_k$ . Closing this gap will be the subject of future study. Additionally, it is planned to carry over our theory to the setting of generalized Nash equilibrium problems, where players' strategy sets are coupled, e.g., over a PDE and additionally physical requirements lead to a constraint on the state.



## A Essentially Bounded and Strongly Measurable Linear Operators

*Proof of Lemma 2.6.* (i) First, notice that for any  $y \in L^\infty(\Omega, Y)$ , we have

$$\|\mathcal{A}(\cdot)y(\cdot)\|_{L^\infty(\Omega, Z)} \leq \|\mathcal{A}\|_{L^\infty(\Omega, \mathcal{L}(Y, Z))} \|y\|_{L^\infty(\Omega, Y)} < \infty.$$

Since the product of strongly measurable functions is again strongly measurable,  $\mathcal{A}(\cdot)y(\cdot) : \Omega \rightarrow Z$  is strongly measurable, meaning  $\mathcal{A}_1 : L^\infty(\Omega, Y) \rightarrow L^\infty(\Omega, Z), y(\cdot) \mapsto \mathcal{A}(\cdot)y(\cdot)$  is well-defined and bounded. Hence the adjoint operator  $\mathcal{A}_1^* : (L^\infty(\Omega, Z))^* \rightarrow (L^\infty(\Omega, Y))^*$  is bounded. Now, let  $w^\circ \in (L^\infty(\Omega, Z))^*$  be a singular functional, i.e., by Definition 2.3 there exists a sequence  $\{F_n\} \subset \mathcal{F}$  with  $F_{n+1} \subset F_n$  such that  $\mathbb{P}(F_n) \rightarrow 0$  and  $\langle w^\circ, z \rangle_{(L^\infty(\Omega, Z))^*, L^\infty(\Omega, Z)} = 0$  for all  $z \in L^\infty(\Omega, Z)$  satisfying  $z(\omega) = 0$  for all  $\omega \in F_n$  (for some  $n$ ). For this same sequence  $\{F_n\}$ , let  $y$  be an arbitrary element of  $L^\infty(\Omega, Y)$  satisfying  $y = 0$  on  $F_n$ . Then linearity of  $\mathcal{A}(\omega)$  implies  $\mathcal{A}(\omega)y(\omega) = 0$  for all  $\omega \in F_n$  and so with  $\mathcal{A}_1 y = \mathcal{A}(\cdot)y(\cdot)$  we have

$$\langle \mathcal{A}_1^* w^\circ, y \rangle_{(L^\infty(\Omega, Y))^*, L^\infty(\Omega, Y)} = \langle w^\circ, \mathcal{A}(\cdot)y(\cdot) \rangle_{(L^\infty(\Omega, Z))^*, L^\infty(\Omega, Z)} = 0.$$

Since  $y$  is arbitrary,  $\mathcal{A}_1^* w^\circ$  is a singular element of  $(L^\infty(\Omega, Y))^*$ . Finally, if  $z^* \in L^1(\Omega, Z^*)$ , we have

$$\|\mathcal{A}^*(\cdot)z^*(\cdot)\|_{L^1(\Omega, Y^*)} \leq \|\mathcal{A}^*\|_{L^\infty(\Omega, \mathcal{L}(Z^*, Y^*))} \|z^*\|_{L^1(\Omega, Z^*)},$$

meaning  $\hat{\mathcal{A}}_1^* : L^1(\Omega, Z^*) \rightarrow L^1(\Omega, Y^*), z^*(\cdot) \mapsto \mathcal{A}^*(\cdot)z^*(\cdot)$  is bounded.

(ii) The second part of the proof follows using analogous arguments to the above.  $\square$

*Proof of Corollary 2.7.* (i) Let  $\{y_n\}$  be a sequence in  $L^\infty(\Omega, Y)$  with  $y_n \xrightarrow{*} \hat{y}$ . We need to show that  $\mathcal{A}_1^* y_n \xrightarrow{*} \mathcal{A}_1^* \hat{y}$ . We have for all  $z^* \in L^1(\Omega, Z^*)$  and all  $n$

$$\begin{aligned} \langle z^*, \mathcal{A}_1 y_n \rangle_{L^1(\Omega, Z^*), L^\infty(\Omega, Z)} &= \mathbb{E}[\langle z^*(\cdot), \mathcal{A}(\cdot)y_n(\cdot) \rangle_{Z^*, Z}] \\ &= \mathbb{E}[\langle \mathcal{A}^*(\cdot)z^*(\cdot), y_n(\cdot) \rangle_{Y^*, Y}] = \langle \hat{\mathcal{A}}_1^* z^*, y_n \rangle_{L^1(\Omega, Y^*), L^\infty(\Omega, Y)} \end{aligned}$$

Hence as  $n \rightarrow \infty$ ,

$$\langle z^*, \mathcal{A}_1 y_n \rangle_{L^1(\Omega, Z^*), L^\infty(\Omega, Z)} \rightarrow \langle \hat{\mathcal{A}}_1^* z^*, \hat{y} \rangle_{L^1(\Omega, Y^*), L^\infty(\Omega, Y)} = \langle z^*, \mathcal{A}_1 \hat{y} \rangle_{L^1(\Omega, Z^*), L^\infty(\Omega, Z)}. \quad (50)$$

Therefore  $\mathcal{A}_1 : L^\infty(\Omega, Y) \rightarrow L^\infty(\Omega, Z)$  is weakly-to-weakly\* continuous. If  $y_n$  belongs to the set  $E := \{y \in L^\infty(\Omega, Y) : \mathcal{A}(\omega)y(\omega) = b(\omega) \text{ a.s.}\}$  for all  $n$ , then (50) gives  $\mathbb{E}[\langle z^*(\cdot), \mathcal{A}(\cdot)\hat{y}(\cdot) \rangle_{Z^*, Z}] = \mathbb{E}[\langle z^*(\cdot), b(\cdot) \rangle_{Z^*, Z}]$  for all  $z^* \in L^1(\Omega, Z^*)$ . Thus  $\mathcal{A}(\omega)\hat{y}(\omega) = b(\omega)$  a.s, meaning  $\hat{y}$  belongs to the set  $E$ , making  $E$  weakly\* closed.

(ii) The second claim follows using similar arguments after noting that  $Y$  is reflexive and taking  $y_n \rightarrow \hat{y}$ .  $\square$

## B Existence of Saddle Points

In this section, we adapt proofs from [19] to include an objective containing the term  $\mathcal{R}[J_2(x)]$  instead of  $\mathbb{E}[J_2(x)]$ . Additionally, we omit the abstract constraint  $x_2(\omega) \in C_2$  a.s.

We define the dual problem

$$\text{maximize}_{\lambda \in \Lambda} \left\{ g(\lambda) := \inf_{x \in X} L(x, \lambda) \right\}. \quad (\text{D})$$

By basic duality, the question of the existence of saddle points is the same as identifying those  $(\bar{x}, \bar{\lambda})$  for which the minimum of Problem (P) and maximum of Problem (D) is attained, i.e.,

$$\inf_{x \in X} \mathbf{P} = \inf_{x \in X} \sup_{\lambda \in \Lambda} L(x, \lambda) = \sup_{\lambda \in \Lambda} \inf_{x \in X} L(x, \lambda) = \sup \text{D}.$$

Let  $\varphi(x, u) = j(x)$  if  $x \in F_{\text{ad}, u}$ , and  $\varphi(x, u) = \infty$  otherwise. This function is related to the Lagrangian by the relationship  $\varphi(x, u) = \sup_{\lambda \in \Lambda} \{L(x, \lambda) - \langle u, \lambda \rangle_{U, \Lambda}\}$ . Additionally, it is related to the value function through  $v(u) = \inf_{x \in X} \varphi(x, u)$ . Obviously,  $v(0) = \inf \text{P}$ .

We proceed by first proving weak\* lower semicontinuity of  $\varphi$ .

**Lemma B.1.** Let Assumption 3.1 be satisfied. Then the function  $\varphi : X \times U \rightarrow \mathbb{R} \cup \{\infty\}$  is weak\* lower semicontinuous.

*Proof.* Let  $X' = X_1^* \times L^1(\Omega, X_2^*)$ , which can be paired with  $X$ . Let  $Y := X \times U$  and denote the pairing on  $Z := X' \times \Lambda$  by

$$\langle y, z \rangle_{Y, Z} := \langle x, x' \rangle_{X, X'} + \langle u, \lambda \rangle_{U, \Lambda}. \quad (51)$$

Since  $Y = Z^*$ , the topology induced by the pairing (51) coincides with the weak\* topology on  $Y$ . We define  $\varphi_1(x) = j(x) + \delta_C(x_1)$  and

$$\varphi_2(x_1, x_2, u; \omega) = \begin{cases} 0, & \text{if } e(x_1, x_2; \omega) = u_e, i(x_1, x_2; \omega) \leq_K u_i, \\ \infty, & \text{otherwise.} \end{cases}$$

Obviously,  $\varphi(x, u) = \varphi_1(x) + \mathbb{E}[\varphi_2(x_1, x_2(\cdot), u(\cdot))]$ . Let  $\langle \cdot, \cdot \rangle_{Y', Z'}$  denote the pairing of  $Y' := X_1 \times X_2 \times (W \times R)$  with  $Z' := X_1^* \times X_2^* \times (W^* \times R^*)$ ; then the conjugate integrand to  $\varphi_2$  is given by

$$\varphi_2^*(z'; \omega) = \sup_{y' \in Y'} \{ \langle y', z' \rangle_{Y', Z'} - \varphi_2(y'; \omega) \}.$$

Defining  $h(y'; \omega) = 0$  for  $y' = (x_1, x_2, u)$  we have  $h(y'; \omega) \leq \varphi_2(y'; \omega)$  a.s. The function  $h$  is trivially a normal convex integrand that is integrable on  $X \times U$ . Thus with the conjugate integrand  $h^*$ ,  $I_h$  and  $I_{h^*}$  are conjugate to each other by Lemma 2.1, meaning that  $I_{h^*} \not\equiv \infty$ .

Since  $h \leq \varphi_2$  we have  $h^* \geq \varphi_2^*$ , and hence there exists a point  $z \in Z$  such that  $I_{\varphi_2^*}(z) < \infty$ . Since there clearly exists a point such that  $I_{\varphi_2}$  is finite, it follows that  $I_{\varphi_2}$  and  $I_{\varphi_2^*}$  are conjugate to one another and are weak\* lower semicontinuous; see Remark 2.2.

Notice that  $J_1 + \delta_C$  is weak\* lower semicontinuous, since it is weakly lower semicontinuous with respect to the natural pairing on the reflexive space  $X_1$ . Hence, together with Proposition 3.4, we have weak\* lower semicontinuity of  $\varphi_1$ .  $\square$

**Theorem B.2.** Let Assumption 3.1 and Assumption 3.5(i) be satisfied. Then

$$-\infty < \min \text{P} = \sup \text{D},$$

meaning that the primal problem (P) attains its minimum, and the minimal value coincides with the supremum of the dual problem (D), which need not be attained.

*Proof.* If  $F_{\text{ad},u}$  is bounded for all  $u$  in a neighborhood of zero, then there exist closed, convex, and bounded sets  $C_1 \subset X_1$  and  $C_2 \subset X_2$  such that if  $x \in F_{\text{ad},u}$ , then  $x_1 \in C_1$  and  $x_2(\omega) \in C_2$  a.s. Then the statement follows by [19, Theorem 3.4]. Otherwise if  $j$  is radially unbounded, then the statement follows by [19, Corollary 3.5]. In both cases, we take advantage of the weak\* lower semicontinuity of  $\varphi$  as proven in Lemma B.1.  $\square$

We now define the extended dual problem by

$$\text{maximize}_{(\lambda, \lambda^\circ) \in \Lambda \times \Lambda^\circ} \left\{ \bar{g}(\lambda, \lambda^\circ) := \inf_{x \in X} \bar{L}(x, \lambda, \lambda^\circ) \right\}. \quad (\bar{D})$$

Clearly,  $\bar{g}(\lambda, 0) = g(\lambda)$  and thus  $\sup D \leq \sup \bar{D}$ . Additionally,  $\sup \bar{D} \leq \inf P$ , since by (15), we have

$$\begin{aligned} \sup_{(\lambda, \lambda^\circ) \in \Lambda \times \Lambda^\circ} \bar{g}(\lambda, \lambda^\circ) &= \sup_{(\lambda, \lambda^\circ) \in \Lambda \times \Lambda^\circ} \inf_{x \in X} \{L(x, \lambda) + L^\circ(x, \lambda^\circ)\} \\ &\leq \inf_{x \in X} \sup_{(\lambda, \lambda^\circ) \in \Lambda \times \Lambda^\circ} \{L(x, \lambda) + L^\circ(x, \lambda^\circ)\}. \end{aligned}$$

**Theorem B.3.** Let Assumption 3.1 and Assumption 3.5(ii) be satisfied. Then

$$\inf P = \max \bar{D} < \infty, \quad (52)$$

meaning that the dual problem ( $\bar{D}$ ) attains its maximum and the maximum value coincides with the infimum of the primal problem (P), which needs not be attained.

*Proof.* Assumption 3.5(ii) guarantees that  $v$  is bounded above on a neighborhood of zero, so (52) follows by, e.g., [34, Theorem 17].  $\square$

**Theorem B.4.** Let Assumption 3.1, Assumption 3.5(ii), and Assumption 3.6 be satisfied. Then

$$\inf P = \max D < \infty,$$

meaning that the dual problem (D) attains its maximum and the maximum value coincides with the infimum of the primal problem (P), which needs not be attained.

*Proof.* This proof follows exactly with the same arguments used in [19, Theorem 3.8]. We repeat them here to keep the paper self-contained. Assumption 3.6 implies

$$\bar{g}(\lambda, \lambda^\circ) \leq g(\lambda) \quad \forall (\lambda, \lambda^\circ) \in \Lambda_0 \times \Lambda_0^\circ. \quad (53)$$

Then in combination with Theorem B.3, this implies

$$\max \bar{D} \leq \sup D \leq \max \bar{D},$$

and the proof is done since a solution  $(\lambda, \lambda^\circ)$  of ( $\bar{D}$ ) gives a solution  $\lambda$  of (D).

To show (53), let  $(\lambda, \lambda^\circ) \in \Lambda_0 \times \Lambda_0^\circ$  be arbitrary. We skip the trivial case  $\bar{g}(\lambda, \lambda^\circ) = -\infty$  and obtain by Lemma 3.8 that

$$\bar{g}(\lambda, \lambda^\circ) = \inf_{x \in X_0} \{L(x, \lambda) + \ell(x_1, \lambda^\circ)\}. \quad (54)$$

We now define

$$h(x_1) = \begin{cases} \inf_{x_2 \in L^\infty(\Omega, X_2)} L(x, \lambda), & \text{if } x_1 \in C, \\ \infty, & \text{else} \end{cases}$$

and

$$k(x_1) = -\ell(x_1, \lambda^\circ).$$

Notice that  $\bar{g}(\lambda, \lambda^\circ) = \inf_{x_1 \in X_1} \{h(x_1) - k(x_1)\}$ . Additionally,  $h \not\equiv \infty$  is convex and  $k > -\infty$  is concave. In fact, since  $\bar{g}$  is finite,  $k$  cannot be identical to  $\infty$  and  $h$  must be proper. Therefore by Fenchel's duality theorem (cf. [5, Theorem 6.5.6]), with  $h^*(v) = \sup_{x_1 \in X_1} \{\langle v, x_1 \rangle_{X_1^*, X_1} - h(x_1)\}$  and  $k^*(v) = \inf_{x_1 \in X_1} \{\langle v, x_1 \rangle_{X_1^*, X_1} - k(x_1)\}$ , we have

$$\bar{g}(\lambda, \lambda^\circ) = \max_{x_1^* \in X_1^*} \{k^*(x_1^*) - h^*(x_1^*)\}. \quad (55)$$

Let  $x_1^*$  denote the maximizer of (55), meaning  $\bar{g}(\lambda, \lambda^\circ) = k^*(x_1^*) - h^*(x_1^*)$ . Then by definition of  $h^*$ , we have for all  $x_1 \in X_1$  that

$$h(x_1) - \langle x_1^*, x_1 \rangle_{X_1^*, X_1} \geq \bar{g}(\lambda, \lambda^\circ) - k^*(x_1^*). \quad (56)$$

Likewise by definition of  $k$  and  $k^*$ , we get

$$\ell(x_1, \lambda^\circ) + \langle x_1^*, x_1 \rangle_{X_1^*, X_1} \geq k^*(x_1^*).$$

It is straightforward to see that  $\ell(x_1, \lambda^\circ) \leq 0$  for all  $x_1 \in \tilde{C}$ . Indeed,  $x_1 \in \tilde{C}$  implies that there exists a  $x_2 \in L^\infty(\Omega, X_2)$  satisfying  $e(x_1, x_2(\omega), \omega) = 0$  and  $i(x_1, x_2(\omega), \omega) \leq_K 0$  a.s. In particular  $L^\circ(x, \lambda^\circ) \leq 0$  and we get

$$\langle x_1^*, x_1 \rangle_{X_1^*, X_1} \geq k^*(x_1^*)$$

for all  $x_1 \in \tilde{C} \supset C$ . From (56) we thus have for all  $x_1 \in C$  that  $h(x_1) \geq \bar{g}(\lambda, \lambda^\circ)$  holds, and hence

$$L(x, \lambda) \geq h(x_1) \geq \bar{g}(\lambda, \lambda^\circ)$$

for all  $x \in X_0$  and all  $(\lambda, \lambda^\circ) \in \Lambda \times \Lambda^\circ$ . It follows that  $g(\lambda) \geq \inf_{x \in X_0} L(x, \lambda) \geq \bar{g}(\lambda, \lambda^\circ)$  and we have shown (53) finishing the proof.  $\square$

## C Additional Proof

*Proof of Lemma 4.1.* By  $K$ -convexity of  $f$ , there exists  $k \in K$  such that for all  $\lambda \in [0, 1]$  and all  $y, \hat{y} \in Y$

$$-f(\lambda y + (1 - \lambda)\hat{y}) = k - \lambda f(y) - (1 - \lambda)f(\hat{y}).$$

Therefore

$$\begin{aligned} g(-f(\lambda y + (1 - \lambda)\hat{y})) &= g(k - \lambda f(y) - (1 - \lambda)f(\hat{y})) \\ &\leq g(k) + g(-\lambda f(y) - (1 - \lambda)f(\hat{y})) \\ &\leq \lambda g(-f(y)) + (1 - \lambda)g(-f(\hat{y})), \end{aligned}$$

where we used convexity of  $g$  for both inequalities.  $\square$

## References

- [1] Robert Adams and John Fournier. *Sobolev Spaces*, volume 140. Elsevier, 2003. doi: 10.1016/S0079-8169(03)80002-8.
- [2] Alen Alexanderian, Noemi Petra, Georg Stadler, and Omar Ghattas. Mean-variance risk-averse optimal control of systems governed by PDEs with random parameter fields using quadratic approximations. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1):1166–1192, 2017. doi: 10.1137/16m106306x.
- [3] Jürgen Appell and Petr P Zabrejko. *Nonlinear superposition operators*, volume 95. Cambridge University Press, 1990. doi: 10.1017/cbo9780511897450.
- [4] Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath. Coherent measures of risk. *Mathematical finance*, 9(3):203–228, 1999. doi: 10.1017/cbo9780511615337.007.
- [5] Jean-Pierre Aubin and Hélène Frankowska. *Set-valued analysis. Modern Birkhäuser Classics*. Birkhäuser Boston Inc., Boston, MA, 1990. doi: 10.1007/978-0-8176-4848-0.
- [6] John M Ball and F Murat. Remarks on Chacon’s biting lemma. *Proceedings of the American Mathematical Society*, 107(3):655–663, 1989. doi: 10.1090/s0002-9939-1989-0984807-3.
- [7] Heinz H Bauschke, Patrick L Combettes, et al. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, volume 408. Springer, 2011. doi: 10.1007/978-1-4419-9467-7.
- [8] Peter Benner, Akwum Onwunta, and Martin Stoll. Block-diagonal preconditioning for optimal control problems constrained by PDEs with uncertain inputs. *SIAM Journal on Matrix Analysis and Applications*, 37(2):491–518, 2016. doi: 10.1137/15m1018502.
- [9] Maïtine Bergounioux and Fredi Tröltzsch. Optimal control of linear bottleneck problems. *ESAIM: Control, Optimisation and Calculus of Variations*, 3:235–250, 1998. doi: 10.1051/cocv:1998108.
- [10] J Frédéric Bonnans and Alexander Shapiro. *Perturbation analysis of optimization problems*. Springer-Verlag, New York, 2013. doi: 10.1007/978-1-4612-1394-9.
- [11] Eduardo Casas. Control of an elliptic problem with pointwise state constraints. *SIAM Journal on Control and Optimization*, 24(6):1309–1318, 1986. doi: 10.1137/0324078.
- [12] Peng Chen and Omar Ghattas. Taylor approximation for chance constrained optimization problems governed by partial differential equations with high-dimensional random parameters. *arXiv preprint arXiv:2011.09985*, 2020.
- [13] Peng Chen, Alfio Quarteroni, and Gianluigi Rozza. A weighted reduced basis method for elliptic partial differential equations with random input data. *SIAM Journal on Numerical Analysis*, 51(6): 3163–3185, 2013. doi: 10.1137/130905253.
- [14] Sergio Conti, Harald Held, Martin Pach, Martin Rumpf, and Rüdiger Schultz. Shape optimization under uncertainty—a stochastic programming perspective. *SIAM Journal on Optimization*, 19(4): 1610–1632, 2008. ISSN 1052-6234. doi: 10.1137/070702059.
- [15] Yu Ermoliev, T Ermolieva, T Kahil, M Obersteiner, V Gorbachuk, and P Knopov. Stochastic optimization models for risk-based reservoir management. *Cybernetics and Systems Analysis*, 55(1):55–64, 2019. doi: 10.1007/s10559-019-00112-z.

- [16] M Hassan Farshbaf-Shaker, René Henrion, and Dietmar Hömberg. Properties of chance constraints in infinite dimensions with an application to PDE constrained optimization. *Set-Valued and Variational Analysis*, 26(4):821–841, 2018. doi: 10.1007/s11228-017-0452-5.
- [17] Deborah B. Gahururu, Michael Hintermüller, and Thomas M. Surowiec. Risk-neutral PDE-constrained generalized Nash equilibrium problems. Preprint SPP1962-155, 2 2021.
- [18] Caroline Geiersbach and Winnifried Wollner. A stochastic gradient method with mesh refinement for pde-constrained optimization under uncertainty. *SIAM Journal on Scientific Computing*, 42(5): A2750–A2772, 2020. doi: 10.1137/19M1263297.
- [19] Caroline Geiersbach and Winnifried Wollner. Optimality conditions for convex stochastic optimization problems in Banach spaces with almost sure state constraints. *Accepted in SIAM Journal on Optimization*, 2021.
- [20] Abebe Geletu, Armin Hoffmann, Patrick Schmidt, and Pu Li. Chance constrained optimization of elliptic PDE systems with a smoothing convex approximation. *ESAIM: Control, Optimisation and Calculus of Variations*, 26:70, 2020. doi: 10.1051/cocv/2019077.
- [21] Michael Hintermüller and Karl Kunisch. Path-following methods for a class of constrained minimization problems in function space. *SIAM Journal on Optimization*, 17(1):159–187, 2006. doi: 10.1137/040611598.
- [22] Michael Hintermüller and Karl Kunisch. Feasible and noninterior path-following in constrained minimization with low multiplier regularity. *SIAM Journal on Control and Optimization*, 45(4): 1198–1221, 2006. doi: 10.1137/050637480.
- [23] Aleksandr Davidovich Ioffe and Vladimir L’vovich Levin. Subdifferentials of convex functions. *Trudy Moskovskogo Matematicheskogo Obshchestva*, 26:3–73, 1972.
- [24] Philip Kolvenbach, Oliver Lass, and Stefan Ulbrich. An approach for robust PDE-constrained optimization with application to shape optimization of electrical engines and of dynamic elastic structures under uncertainty. *Optimization and Engineering*, 19(3):697–731, 2018. doi: 10.1007/s11081-018-9388-3.
- [25] Drew Philip Kouri and Thomas M Surowiec. Risk-averse PDE-constrained optimization using the conditional value-at-risk. *SIAM Journal on Optimization*, 26(1):365–396, 2016. doi: 10.1051/cocv/2019061.
- [26] Drew Philip Kouri and Thomas M Surowiec. Existence and optimality conditions for risk-averse PDE-constrained optimization. *SIAM/ASA Journal on Uncertainty Quantification*, 6(2):787–815, 2018. doi: 10.1137/16M1086613.
- [27] Drew Philip Kouri, Matthias Heinkenschloss, Denis Ridzal, and Bart G van Bloemen Waanders. A trust-region algorithm with adaptive stochastic collocation for PDE optimization under uncertainty. *SIAM Journal on Scientific Computing*, 35(4):A1847–A1879, 2013. doi: 10.1137/120892362.
- [28] Vladimir L’vovich Levin. The Lebesgue decomposition for functionals on the vector-function space  $L_X^\infty$ . *Functional Analysis and Its Applications*, 8(4):314–317, 1974. doi: 10.1007/bf01075488.
- [29] John M Mulvey and Bala Shetty. Financial planning via multi-stage stochastic optimization. *Computers & Operations Research*, 31(1):1–20, 2004. doi: 10.1016/s0305-0548(02)00141-7.

- [30] Mario VF Pereira and Leontina MVG Pinto. Multi-stage stochastic optimization applied to energy planning. *Mathematical programming*, 52(1):359–375, 1991. doi: 10.1007/bf01582895.
- [31] Georg Pflug and Alois Pichler. *Multistage Stochastic Optimization*. Springer, 2014. doi: 10.1007/978-3-319-08843-3.
- [32] Georg Pflug and Werner Römisch. *Modeling, Measuring and Managing Risk*. World Scientific, 2007. doi: 10.1142/6478.
- [33] Ralph Tyrrell Rockafellar. Convex integral functionals and duality. In *Contributions to nonlinear functional analysis*, pages 215–236. Elsevier, 1971. doi: 10.1016/b978-0-12-775850-3.50012-1.
- [34] Ralph Tyrrell Rockafellar. *Conjugate duality and optimization*. SIAM, 1974. doi: 10.1137/1.9781611970524.
- [35] Ralph Tyrrell Rockafellar and Johannes Royset. Engineering decisions under risk averseness. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, 1(2):04015003, 2015. doi: 10.1061/AJRUA6.0000816.
- [36] Ralph Tyrrell Rockafellar and Roger J-B Wets. Stochastic convex programming: Kuhn–Tucker conditions. *Journal of Mathematical Economics*, 2(3):349–370, 1975. doi: 10.1016/0304-4068(75)90003-8.
- [37] Ralph Tyrrell Rockafellar and Roger J-B Wets. Stochastic convex programming: basic duality. *Pacific Journal of Mathematics*, 62(1):173–195, 1976. URL <http://projecteuclid.org/euclid.pjm/1102867873>.
- [38] Ralph Tyrrell Rockafellar and Roger J-B Wets. Stochastic convex programming: singular multipliers and extended duality singular multipliers and duality. *Pacific Journal of Mathematics*, 62(2):507–522, 1976. URL <http://projecteuclid.org/euclid.pjm/1102867735>.
- [39] Ralph Tyrrell Rockafellar and Roger J-B Wets. Stochastic convex programming: relatively complete recourse and induced feasibility. *SIAM Journal on Control and Optimization*, 14(3):574–589, 1976. doi: 10.1137/0314038.
- [40] Claudia Schillings, Stephan Schmidt, and Volker Schulz. Efficient shape optimization for certain and uncertain aerodynamic design. *Computers & Fluids*, 46(1):78–87, 2011. doi: 10.1016/j.compfluid.2010.12.007.
- [41] Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory*. SIAM, Philadelphia, 2009. doi: 10.1137/1.9780898718751.
- [42] JJ Uhl. Extensions and decompositions of vector measures. *Journal of the London Mathematical Society*, 2(4):672–676, 1971. doi: 10.1112/jlms/s2-3.4.672.