AN INTERIOR-POINT APPROACH FOR SOLVING RISK-AVERSE PDE-CONSTRAINED OPTIMIZATION PROBLEMS WITH COHERENT RISK MEASURES

S. GARREIS*, T. M. SUROWIEC†, AND M. ULBRICH‡

Abstract. The prevalence of uncertainty in models of engineering and the natural sciences necessitate the inclusion of random parameters in the underlying partial differential equations (PDEs). The resulting decision problems governed by the solution of such random PDEs are infinite dimensional stochastic optimization problems. In order to obtain risk-averse optimal decisions in the face of such uncertainty, it is common to employ risk measures in the objective function. This leads to risk-averse PDE-constrained optimization problems. We propose a method for solving such problems in which the risk measures are convex combinations of the mean and conditional value-at-risk (CVaR). Since these risk measures can be evaluated by solving a related inequality-constrained optimization problem, we suggest a log-barrier technique to approximate the risk measure. This leads to a new continuously differentiable convex risk measure: the log-barrier risk measure. We show that the log-barrier risk measure fits into the setting of optimized certainty equivalents of Ben-Tal and Teboulle and the expectation quadrangle of Rockafellar and Uryasev. Using the differentiability of the log-barrier risk measure, we derive first-order optimality conditions reminiscent of classical primal and primal-dual interior point approaches in nonlinear programming. We derive the associated Newton system, propose a reduced symmetric system to calculate the steps, and provide a sufficient condition for local superlinear convergence in the continuous setting. Furthermore, we provide a Γ-convergence result for the log-barrier risk measures to prove convergence of the minimizers to the original nonsmooth problem. The results are illustrated by a numerical study.

Key words. Risk-Averse, PDE-Constrained Optimization, Risk Measures, Uncertainty Quantification, Stochastic Optimization, Interior-Point Methods, Conditional Value-at-Risk, Gamma Convergence

AMS subject classifications. 49J20, 49J50, 49J55, 49K20, 49K45, 90C15.

1. Introduction.

Uncertainty is an unavoidable component of practically every complex or data-driven system arising in engineering and the natural sciences. For example, we encounter uncertainty as a result of noisy data measurements, unknown operating parameters, or even unclear assumptions in the modeling of subsurface flows [18, 49], plate tectonics and ice sheet models [35, 48], and next-generation aeronautics designs [8]. In the context of optimization and optimal control, we are tasked with optimizing constrained systems of partial differential equations (PDEs) laden with uncertain inputs, which may arise in the coefficients as well as the bulk and boundary data. This has led to a growing interest in stochastic PDE-constrained optimization.

Whenever we are faced with making a decision under uncertainty, it is important to obtain optimal designs, decisions, or controls that are somehow robust or risk-averse to risky or tail events. Despite having been primarily developed for finite dimensional optimization problems, the stochastic programming literature offers a number of approaches for risk-averse decision-making under uncertainty, see, e.g., [7, 36, 46] and the many references therein. For instance, one might try to solve a minimization problem with stochastic order constraints based on a benchmark design as in probabilistic programming, e.g., [16, 37]. Another approach would be to minimize a kind of worst-case expectation of the quantity of interest or objective function

*Technische Universität München (garreis@ma.tum.de)

†Philipps-Universität Marburg (surowiec@mathematik.uni-marburg.de) TMS's research was sponsored by DFG grant no. SU 963/1-1 "Generalized Nash Equilibrium Problems with Partial Differential Operators: Theory, Algorithms, and Risk Aversion".

‡Chair of Mathematical Optimization, Technische Universität München (mulbrich@ma.tum.de)

over a class of probability measures as in distributionally robust optimization, e.g., [17, 45, 47]. Yet another possibility is to employ risk measures, see, e.g., [38, 42, 43] as well as [46, Chap. 6], which allows a broad degree of flexibility and yields structures that may be more familiar to researchers working in PDE-constrained optimization or optimal control.

In this paper, we take the latter approach and follow the framework developed in [30, 32] for risk-averse PDE-constrained optimization. Thus, we consider the following abstract infinite dimensional stochastic optimization problem:

$$\min_{z \in \mathcal{Z}_{\mathrm{ad}}} \mathcal{R}[\mathcal{J}(S(z))] + \wp(z), \tag{1.1}$$

where $z \in Z$ are deterministic decisions (designs, controls, etc.), $\mathcal{Z}_{\mathrm{ad}}$ is the admissible set, $\wp$ is a function modeling the cost of z, $\mathcal{R}$ is a functional that maps a set of random variables $X \in \mathcal{X}$ into $\overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$ called a risk measure, and $\mathcal{J}$ is an uncertain objective function, quantity of interest, or cost that depends on the z-dependent solution of the PDE with uncertain inputs, denoted throughout by $S(z)$. Note that $S(z)$ itself is a random field. We use $(\Omega, \mathcal{F}, \mathbb{P})$ to denote a probability space and the expectation by $\mathbb{E}$, i.e., if $X : \Omega \to \mathbb{R}$ is a random variable then $\mathbb{E}[X] = \int_\Omega X(\omega)\mathrm{d}\mathbb{P}(\omega)$.

There have been a number of recent contributions to PDE-constrained optimization under uncertainty in theory, algorithms, and numerical approximation schemes, e.g., [14, 41, 50]. However, the overwhelming majority of work on numerical approximation and solution algorithms has been for the risk-neutral case in which $\mathcal{R} = \mathbb{E}$.

The risk-neutral case provides solutions $z^\star$ that perform well on *average*. Therefore, employing such a decision $z^\star$ is only reasonable if a task is to be performed many times over. Despite this, there is still no way of accounting for possibly catastrophic tail events. In contrast, we choose a class of risk measures particularly suited to yield solutions $z^\star$ that mitigate tail risk.

For literature on numerical approximation schemes, we highlight here the work on reduced-order model approaches [12, 13], spatial multigrid algorithms with sparse-grid collocation [10, 11], low-rank tensor approximation [6, 24], and numerical solution and optimization methods based on Taylor expansions [1, 20, 21, 22, 34]. Unlike numerical approximation, the literature is rather scarce on dedicated optimization algorithms for PDE-constrained optimization under uncertainty. In addition to [1, 20, 34], we point to [28, 29] for a globally convergent trust-region algorithm based on adaptive sparse grids. Though the latter was developed for the risk-neutral case with $\mathcal{Z}_{\mathrm{ad}} = Z$, it can be easily extended to include smooth risk measures and bound constraints $\mathcal{Z}_{\mathrm{ad}}$.

Risk-averse PDE-constrained optimization, i.e., where $\mathcal{R}[X] > \mathbb{E}[X]$ for all non-constant random variables X, is much more recent both from a theoretical and algorithmic perspective, see, e.g., [1, 6, 27, 30, 31, 32, 34]. In [30, 31], variational regularization techniques are developed that allow the application of algorithms for smooth PDE-constrained optimization, as mentioned above, whereas [32] presents a general existence and optimality theory. Although [27, 34] take the perspective of *robust* optimization, i.e., $\mathcal{R}[X] = \sup_{\omega \in \Omega} |X(\omega)|$, we mention it here as any solutions obtained using this method would be clearly risk-averse.

The goal of this paper is to develop an interior-point method for the solution of (1.1) when $\mathcal{R}$ (a non-smooth risk measure) is defined by

$$\mathcal{R}[X] := \inf_{t \in \mathbb{R}} \left\{ t + \mathbb{E}[v(X - t)] \right\}, \tag{1.2}$$

where $v : \mathbb{R} \to \mathbb{R}$ is given by

$$v(s) = \max\{a_1 s, a_2 s\}, \; a_1 \in [0,1) \text{ and } a_2 \in (1, \infty). \tag{1.3}$$

Here, v is a so-called scalar regret function (negative utility function) that implies a certain aversion to risk when used in (1.2). Our approach combines modern techniques of interior-point methods for infinite dimensional PDE-constrained optimization problems as in [24, 44, 53, 54] with the available theory of risk-averse PDE-constrained optimization mentioned above.

In particular, the choice of the scalar regret function implies that $\mathcal{R}$ is a so-called coherent risk measure generated by the expectation quadrangle, see [39] as well as the earlier work [4] and [5]. This includes the popular conditional value-at-risk functional CVaR_β (also called average value-at-risk, expected shortfall, tail expectation), which is more intuitively defined as a tail expectation by

$$\mathrm{CVaR}_\beta[X] := \tfrac{1}{1-\beta}\int_\beta^1 F_X^{-1}(\alpha)\mathrm{d}\alpha, \quad \beta \in (0,1),$$

where $F_X^{-1}(\alpha)$ is the α-quantile (value-at-risk) of the random variable X, $\beta := (a_2 - 1)/a_2$, and $a_1 = 0$. In fact, the form of v implies that $\mathcal{R}$ in this paper is any convex combination of the expected value and CVaR, but not the expected value alone.

In light of the assumptions on $\mathcal{R}$, we may rewrite (1.1) in an alternative form by introducing slacks $t \in \mathbb{R}$, $W \in \mathcal{X}$, and two inequality constraints:

(1.4)
$$\min_{(z,W,t)\in\mathcal{Z}_{\mathrm{ad}}\times\mathcal{X}\times\mathbb{R}} t + \mathbb{E}[W] + \wp(z) \ \text{ s.t. } \begin{cases} W \geq a_1(\mathcal{J}(S(z)) - t), & \mathbb{P}\text{-a.a. } \omega \in \Omega, \\ W \geq a_2(\mathcal{J}(S(z)) - t), & \mathbb{P}\text{-a.a. } \omega \in \Omega. \end{cases}$$

This is a commonplace reformulation often used in stochastic programming. Nevertheless, although we have removed the nonsmoothness from the objective, (1.4) retains the complexity introduced by $\mathcal{R}$ due to the potentially non-convex inequality constraints. Moreover, there are no available algorithms for stochastic PDE-constrained optimization problems with nonlinear state constraints; even in this local/global setting where the state $S(z)$ is treated globally in the sense that $\mathcal{J}$ "integrates out" the spacial dependence and locally in that $W - a_i(\mathcal{J}(S(z)) - t) \in \mathcal{X}$ is a random variable. Of course, if S is affine and $\mathcal{J}$ is convex with respect to the usual partial order on $\mathcal{X}$, then these constraints would be convex.

Inspired by the success of interior-point methods for parameteric variational inequalities in [24], we propose an approach in which we (approximately) solve a sequence of μ-dependent ($\mu > 0$) log-barrier-approximations of (1.4) given by

$$(1.5) \qquad \min_{(z,W,t)\in\mathcal{Z}_{\mathrm{ad}}\times\mathcal{X}\times\mathbb{R}} \mathbb{E}\Big[t + W - \mu \sum_{i=1,2} \ln(W - a_i(\mathcal{J}(S(z)) - t))\Big] + \wp(z).$$

Depending on the explicit structure, the subproblems can be solved by either a semismooth Newton method, see e.g., [26, 51, 52], if $\mathcal{Z}_{\mathrm{ad}}$ is defined by simple bound constraints, or a trust-region approach as in [28, 29].

As an added bonus of the proposed optimization method, we obtain a new class of risk measures, which we refer to as "log-barrier" risk measures. The log-barrier risk functionals can be shown to arise from the expectation quadrangle for a certain choice of scalar regret function v. This allows us to analyze the associated optimization problems by leveraging the analysis in [32]. For instance, we obtain familiar primal and primal-dual optimality systems as in traditional interior-point approaches. Furthemore, the log-barrier risk measures are amenable to either a traditional sample-based Monte-Carlo approximation, cf. [46, Chap. 5], or the low-rank tensor approximation developed in [24].

The rest of the paper is structured as follows. In Section 2 we provide the necessary notation and data assumptions. Afterwards, in Section 3 we analyze the log-barrier risk measure $\mathcal{R}_\mu$. Then in Section 4, we summarize several important results from the literature and prove the existence of a solution to the approximating problems and derive associated optimality conditions. We show that the optimality conditions can be rewritten as a purely primal or primal-dual system reminiscent of classical interior-point methods. In Section 5, we provide a thorough analysis of the Newton system for the continuous, i.e., function-space setting. This is followed by an asymptotic analysis of $\mathcal{R}_\mu$ as $\mu \downarrow 0$ in Section 6, where we employ several results and ideas from the theory of Γ-convergence to finally demonstrate the convergence of minimizers as $\mu \downarrow 0$. Finally, in Section 7, we demonstrate the viability of the approach by solving a model problem numerically.

2. Notation, Assumptions, and Preliminary Results.

2.1. Spaces of Random Variables. Let Ω be a non-empty set and $\mathcal{F}$ an associated σ-algebra. Throughout the text, $(\Omega, \mathcal{F}, \mathbb{P})$ denotes a complete probability space, where the set-function $\mathbb{P} : \mathcal{F} \to [0, 1]$ is a probability measure. Whenever it is clear in context, we use "a.e." and "a.a." to denote "almost everywhere" and "almost all", respectively. Furthermore, we denote the expectation of some random variable $X : \Omega \to \mathbb{R}$ by $\mathbb{E}[X] := \int_\Omega X(\omega) \mathrm{d}\mathbb{P}(\omega)$.

We will make assumptions below that require the random quantities to have a certain degree of integrability. Therefore, we make use of Bochner spaces to characterize the random quantities. We recall that the Bochner space $L^p(\Omega, \mathcal{F}, \mathbb{P}; W)$ comprises all strongly measurable functions from $(\Omega, \mathcal{F}, \mathbb{P})$ into some Banach space W with p finite absolute moments $p \in [1, \infty)$. $L^\infty(\Omega, \mathcal{F}, \mathbb{P}; W)$ is the space of $\mathbb{P}$-essentially bounded W-valued strongly measurable functions, cf. [25] for a full discussion. When endowed with the corresponding norm given by:

$$\|v\|_{L^p(\Omega,\mathcal{F},\mathbb{P};W)} = \mathbb{E}\left[\|v\|_W^p\right]^{1/p} \text{ for } p \in [1, \infty) \quad \text{or} \quad \|v\|_{L^\infty(\Omega,\mathcal{F},\mathbb{P};W)} = \operatorname*{ess\,sup}_{\omega \in \Omega} \|v(\omega)\|_W$$

$L^p(\Omega, \mathcal{F}, \mathbb{P}; W)$ is a Banach space. As is commonly the case, we use the convention

$$L^p(\Omega, \mathcal{F}, \mathbb{P}) = L^p(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}),$$

whenever $W = \mathbb{R}$. For readability, we will often use the simplifying notation

$$\mathcal{X} := L^p(\Omega, \mathcal{F}, \mathbb{P}).$$

In addition, if $p = 1$, then we identify $\mathcal{X}^* = (L^1(\Omega, \mathcal{F}, \mathbb{P}))^*$ with $L^\infty(\Omega, \mathcal{F}, \mathbb{P})$.

2.2. General Spaces. We assume that the deterministic decision space Z and solution space U are real reflexive Banach spaces. The associated feasible/admissible set of decisions is denoted by $\mathcal{Z}_{\text{ad}} \subset Z$ and is assumed to be nonempty, closed, and convex. Given two real Banach spaces X and Y, we denote the space of bounded linear operators from X into Y by $\mathcal{L}(X, Y)$. Of course, if $Y = \mathbb{R}$, then $X^* := \mathcal{L}(X, \mathbb{R})$ denotes the topological (continuous) dual space of X and $\langle \cdot, \cdot \rangle_{X^*,X}$ denotes the associated duality pairing. For some bounded linear operator $A \in \mathcal{L}(X, Y)$ we denote by $A^* \in \mathcal{L}(Y^*, X^*)$ the adjoint (dual, conjugate) operator of A. Strong convergence (w.r.t. the norm topology) is denoted by "$\to$", whereas "$\rightharpoonup$" denotes weak and "$\overset{*}{\rightharpoonup}$" weak* convergence.

2.3. Risk Measures. As mentioned in the introduction, we assume that the risk measure $\mathcal{R}$ given in (1.2) is generated by a scalar regret (negative utility) function $v : \mathbb{R} \to \mathbb{R}$. This goes back to an idea of Ben-Tal and Teboulle [4, 5], see also [39], in which the term *optimized certainty equivalent* (OCE) was used. We provide the following result as a summary of the basic construction and properties of the associated risk measure.

THEOREM 2.1. *Let* $\mathcal{X} := L^1(\Omega, \mathcal{F}, \mathbb{P})$ *and let* $v : \mathbb{R} \to \overline{\mathbb{R}}$ *be closed, convex and increasing such that*

$$v(0) = 0 \quad \textit{and} \quad v(x) > x \;\forall\, x \neq 0. \tag{2.1}$$

For $X \in \mathcal{X}$*, suppose that* $\mathcal{V}(X) := \mathbb{E}[v(X)]$ *and define* $\mathcal{R} : \mathcal{X} \to \overline{\mathbb{R}}$ *by*

$$\mathcal{R}(X) := \inf_{t\in\mathbb{R}} \{t + \mathcal{V}(X - t)\}.$$

Then $\mathcal{V} : \mathcal{X} \to \overline{\mathbb{R}}$ *is proper, closed, convex, and fulfills*

$$\mathcal{V}(X) > \mathbb{E}[X] \quad \forall\, X \neq 0 \;\mathbb{P}\text{-}a.e. \;\textit{and}\; \lim_{k\to\infty} \{\mathcal{V}(X_k) - \mathbb{E}[X_k]\} = 0 \implies \lim_{k\to\infty} \mathbb{E}[X_k] = 0.$$

The statistic, $S(X) \subset \mathbb{R}$*, given by*

$$\mathcal{S}(X) := \operatorname*{argmin}_{t\in\mathbb{R}} \{t + \mathcal{V}(X - t)\}$$

is a non-empty compact interval for any $X \in \mathcal{X}$*. In addition,* $\mathcal{R} : \mathcal{X} \to \overline{\mathbb{R}}$ *is proper, closed, convex, and satisfies*

(2.2a)	(Invariance on Constants) :	$\mathcal{R}(C) = C$ *for all* $C \in \mathbb{R}$.
(2.2b)	(Risk Aversion) :	$\mathcal{R}(X) > \mathbb{E}[X]$ *for all non-constant* $X \in \mathcal{X}$.
(2.2c)	(Translation Equivariance) :	$\mathcal{R}(X + C) = \mathcal{R}(X) + C$ *for all* $C \in \mathbb{R}$.
(2.2d)	(Monotonicity) :	$X \leq X'$ $\mathbb{P}$-*a.a.* $\omega \in \Omega \Longrightarrow \mathcal{R}(X) \leq \mathcal{R}(X')$.

Proof. See [32, Section 2.4] and [32, Appendix] for a rigorous derivation in general Lebesgue spaces. ∎

We will exploit the statement of Theorem 2.1 in our analysis of the log-barrier risk measure. Note also that $\mathcal{R}$ in the previous theorem is a regular measure of risk in the sense of Rockafellar and Uryasev and satisfies three of the four axioms of coherent measures of risk. If, in addition, $\mathcal{R}$ is positively homogeneous, then it is a coherent risk measure, cf. [3].

3. The Log-Barrier Risk Measure. Returning to the discussion leading up to (1.5) in the introduction, we restrict our attention to $\mathcal{R}$ as defined in (1.2), which we restate here for convenience. Let $X \in \mathcal{X}$ (assuming $p = 1$), then

$$\mathcal{R}(X) = \inf_{t\in\mathbb{R}} \mathbb{E}[t + \max\{a_1(X - t), a_2(X - t)\}].$$

Clearly, we can use the same transformation as in (1.4) and obtain

$$\mathcal{R}(X) = \inf_{t\in\mathbb{R}, W\in\mathcal{X}} \left\{\mathbb{E}[t + W] \,|\, W \geq a_i(X - t),\ \mathbb{P}\text{-a.a. } \omega \in \Omega,\ i \in \{1, 2\}\right\}.$$

This is then approximated by the log-barrier risk measure

$$\mathcal{R}_\mu(X) := \inf_{t\in\mathbb{R}, W\in\mathcal{X}} \Big\{ \mathbb{E}\Big[t + W - \mu \sum_{i=1}^{2} \ln(W - a_i(X-t)) \Big] + \zeta(\mu) \Big\}, \quad \mu > 0,$$

where

$$\zeta(\mu) := \mu\big(\ln\big(\tfrac{a_2-a_1}{a_2-1}\mu\big) + \ln\big(\tfrac{a_2-a_1}{1-a_1}\mu\big) - 2\big) \in \mathbb{R} \tag{3.1}$$

is a constant shift needed to ensure that (2.1) holds. This ultimately leads to (1.5). Next, we introduce the functional $F_\mu : \mathcal{X} \times \mathcal{X} \times \mathbb{R} \to \overline{\mathbb{R}}$ given by

$$F_\mu(X, W, t) := \mathbb{E}[t + W - \mu \ln(W - a_1(X-t)) - \mu \ln(W - a_2(X-t)) + \zeta(\mu)].$$

Proposition 3.1. *For every $\mu > 0$ and any $X \in \mathcal{X}$, we have*

$$\inf_{t\in\mathbb{R}, W\in\mathcal{X}} F_\mu(X, W, t) = \inf_{t\in\mathbb{R}} \mathbb{E}[\inf_{w\in\mathbb{R}} f_\mu(X(\cdot), w, t)],$$

where $f_\mu : \mathbb{R} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is defined by

$$f_\mu(x, w, t) := t + w - \mu \ln(w - a_1(x-t)) - \mu \ln(w - a_2(x-t)) + \zeta(\mu). \tag{3.2}$$

for $(x, w, t) \in \mathbb{R}^3$.

Proof. We first observe that

$$\inf_{t\in\mathbb{R}, W\in\mathcal{X}} F_\mu(X, W, t) = \inf_{t\in\mathbb{R}} \inf_{W\in\mathcal{X}} \mathbb{E}[f_\mu(X(\cdot), W(\cdot), t)]. \tag{3.3}$$

Continuing, we will use the theory of normal integrands, cf. [40, Chap. 14], to prove the assertion. To this aim, note that the space $\mathcal{X} = L^1(\Omega, \mathcal{F}, \mathbb{P})$ is decomposable in the sense of [40, Def. 14.59] and, as a probability measure, $\mathbb{P}$ is σ-finite. Next, given some fixed $X \in \mathcal{X}$ and $t \in \mathbb{R}$, we claim that the function $\widehat{f}_\mu : \Omega \times \mathbb{R} \to \overline{\mathbb{R}}$ defined by

$$\widehat{f}_\mu(\cdot, w) := f_\mu(X(\cdot), w, t)$$

is a normal integrand in the sense of [40, Def. 14.27]. Indeed, the mapping

$$\mathbb{R}^2 \ni (x, w) \mapsto f_\mu(x, w, t)$$

is lower semicontinuous and jointly convex (independently of $\omega \in \Omega$). Furthermore, $\operatorname{int} \operatorname{dom}(f_\mu(\cdot, \cdot, t)) \neq \emptyset$ and the mapping $\Omega \ni \omega \mapsto f_\mu(x, w, t)$ is trivially measurable for all $w \in \mathbb{R}$, as $f_\mu(x, w, t)$ is independent of $\omega \in \Omega$. Hence, the mapping

$$\Omega \times \mathbb{R}^2 \ni (\omega, x, w) \mapsto f_\mu(x, w, t)$$

is a normal integrand by [40, Proposition 14.39]. Moreover, the composition rule [40, Prop. 14.45(c)] implies that $\widehat{f}_\mu$ is a normal integrand. Finally, letting

$$\widetilde{W} := \max\{a_1(X-t), a_2(X-t), 0\} + 1 \in \mathcal{X},$$

we see that the critical components of $\widehat{f}_\mu(\cdot, \widetilde{W}(\cdot))$ remain bounded $\mathbb{P}$-a.e. due to the fact that

$$\widetilde{W} - a_i(X-t) \geq 1 \text{ and } \ln(\widetilde{W} - a_i(X-t)) \leq \max\{a_1(X-t), a_2(X-t), 0\} - a_i(X-t).$$

Hence, there exists $\widetilde{W} \in \mathcal{X}$ such that $\int_\Omega \widehat{f}_\mu(\cdot, \widetilde{W}(\cdot))\mathrm{d}\mathbb{P}(\Omega) < \infty$, and we can apply the "interchangeability theorem" [40, Thm. 14.60] to derive

$$\inf_{W \in \mathcal{X}} \mathbb{E}[f_\mu(X(\cdot), W(\cdot), t)] = \inf_{W \in \mathcal{X}} \int_\Omega \widehat{f}_\mu(\omega, W(\omega))\mathrm{d}\mathbb{P}(\omega)$$
$$= \int_\Omega \inf_{w \in \mathbb{R}} \left\{ \widehat{f}_\mu(\omega, w) \right\} \mathrm{d}\mathbb{P}(\omega) = \mathbb{E}[\inf_{w \in \mathbb{R}} f_\mu(X(\cdot), w, t)].$$

This shows the desired result together with (3.3). ∎

In light of Proposition 3.1, we consider for each $t \in \mathbb{R}$, $\mathbb{P}$-a.a. $\omega \in \Omega$, and $x = X(\omega)$ the one-dimensional problem

$$\min_{w \in \mathbb{R}} f_\mu(x, w, t),$$

where the unknown w stands for $W(\omega)$. Due to the explicit structure of f_μ, we can obtain a useful closed formula for the unique optimal solution $\overline{w}$ as a function of x, t. As a result, we will obtain a new scalar regret (negative utility) function v_μ.

Proposition 3.2. *Fix $\mu > 0$, $t \in \mathbb{R}$, $\omega \in \Omega$, and set $x = X(\omega)$. The function $\mathbb{R} \ni w \mapsto f_\mu(x, w, t)$ with f_μ defined in (3.2) has the unique minimizer*

$$\overline{w} = w_\mu(x - t) := \mu + \tfrac{(a_1+a_2)(x-t)+\sqrt{(a_2-a_1)^2(x-t)^2+4\mu^2}}{2}. \tag{3.4}$$

Proof. Let $\varphi(\cdot) := f_\mu(x, \cdot, t)$. Clearly, $\varphi : \mathbb{R} \to \mathbb{R}$ is finite, convex, and continuously differentiable provided $w > a_i(x - t)$ $(i \in \{1, 2\})$, where

$$\varphi'(w) = 1 - \tfrac{\mu}{w - a_1(x-t)} - \tfrac{\mu}{w - a_2(x-t)}.$$

After some elementary computations, we see that the equation $\varphi'(w) = 0$ has one root given by $\overline{w}$ in (3.4); whereas the other root

$$\mu + \tfrac{1}{2}\left((a_1 + a_2)(x - t) - \sqrt{(a_2 - a_1)^2(x - t)^2 + 4\mu^2}\right)$$

would violate the feasibility requirement that $w > a_i(x - t)$, i.e., the objective would be equal to $+\infty$. The assertion follows. ∎

In order to prove that $\mathcal{R}_\mu$ is generated by the expectation quadrangle/as an optimized certainty equivalent, we will need the following short technical lemma.

Lemma 3.3. *Let $w : \mathbb{R} \to \mathbb{R}$ be a given function and $d \in \mathbb{R}$ a constant. Then the function $\widehat{w} : \mathbb{R} \to \mathbb{R}$, $\widehat{w}(s) := w(s + d) - d$ induces the same risk measure as w.*

Proof. Fix $X \in \mathcal{X}$ and observe that

$$\inf_{t \in \mathbb{R}} \{t + \mathbb{E}[\widehat{w}(X - t)]\} = \inf_{t \in \mathbb{R}} \{t + \mathbb{E}[w(X - t + d) - d]\} \stackrel{\tilde{t}=t-d}{=} \inf_{\tilde{t} \in \mathbb{R}} \{\tilde{t} + \mathbb{E}[w(X - \tilde{t})]\}.$$

∎

Proposition 3.4. *For every $\mu > 0$ and any $X \in \mathcal{X}$, we have*

$$\mathcal{R}_\mu(X) = \inf_{t \in \mathbb{R}} \{t + \mathbb{E}[v_\mu(X - t)]\},$$

where

$$v_\mu(s) := w_\mu(s) - \mu \ln(w_\mu(s) - a_1 s) - \mu \ln(w_\mu(s) - a_2 s) + \zeta(\mu) \tag{3.5}$$

and w_μ is given by (3.4).

Proof. By Proposition 3.1, we have

$$\mathcal{R}_\mu(X) = \inf_{t\in\mathbb{R}, W\in\mathcal{X}} F_\mu(X,W,t) = \inf_{t\in\mathbb{R}} \mathbb{E}[\inf_{w\in\mathbb{R}} f_\mu(X(\cdot),w,t)].$$

Then using (3.4), we obtain

$$\mathcal{R}_\mu(X) = \inf_{t\in\mathbb{R}} \mathbb{E}[f_\mu(X(\cdot),\overline{W}_{X,t}(\cdot),t)],$$

where for $\mathbb{P}$-a.a. $\omega \in \Omega$

$$\overline{W}_{X,t}(\omega) := \mu + \tfrac{1}{2}\big((a_1+a_2)(X(\omega)-t) + \sqrt{(a_2-a_1)^2(X(\omega)-t)^2+4\mu^2}\big). \tag{3.6}$$

Substituting this formula into the previous relation yields the assertion. ∎

In our next result, we prove the necessary properties of the new scalar regret function v_μ that allow us to apply the results of Subsection 2.3, along with those of Subsection 4.2 below, to $\mathcal{R}_\mu$ and the associated optimization problems.

Proposition 3.5. *For any $\mu > 0$, the scalar regret function $v_\mu : \mathbb{R} \to \mathbb{R}$ is twice continuously differentiable, strictly convex, strictly increasing. In addition, we have*

$$|v_\mu(s) - v_\mu(s')| \le a_2\,|s-s'|, \quad \forall s, s' \in \mathbb{R}. \tag{3.7}$$

Proof. Let $s \in \mathbb{R}$. Using basic calculus techniques, one can show after some computation, cf. [23], that $v_\mu \in C^2(\mathbb{R})$ with derivatives

$$v'_\mu(s) = w'_\mu(s) - \mu\,\tfrac{w'_\mu(s)-a_1}{w_\mu(s)-a_1 s} - \mu\,\tfrac{w'_\mu(s)-a_2}{w_\mu(s)-a_2 s}, \tag{3.8}$$

$$v''_\mu(s) = \tfrac{\mu(a_2-a_1)^2}{2\mu\sqrt{(a_2-a_1)^2s^2+4\mu^2}+(a_2-a_1)^2s^2+4\mu^2}, \tag{3.9}$$

where

$$w'_\mu(s) = \tfrac{a_1+a_2}{2} + \tfrac{(a_2-a_1)^2 s}{2\sqrt{(a_2-a_1)^2s^2+4\mu^2}}.$$

Therefore, v_μ is proper and, since $v''_\mu(s) > 0$ for all $s \in \mathbb{R}$, v_μ is strictly convex.

Turning to monotonicity, since $a_2 > a_1$ we have

$$\lim_{s\to-\infty} w'_\mu(s) = \tfrac{a_1+a_2}{2} - \tfrac{|a_2-a_1|}{2} = a_1 \text{ and } \lim_{s\to+\infty} w'_\mu(s) = \tfrac{a_1+a_2}{2} + \tfrac{|a_2-a_1|}{2} = a_2.$$

Moreover,

$$\tfrac{w'_\mu(s)-a_1}{w_\mu(s)-a_1 s} = \frac{\tfrac{a_2-a_1}{2} + \tfrac{(a_2-a_1)^2 s}{2\sqrt{(a_2-a_1)^2s^2+4\mu^2}}}{\mu + \tfrac{(a_2-a_1)s}{2} + \tfrac{\sqrt{(a_2-a_1)^2s^2+4\mu^2}}{2}} \longrightarrow 0 \quad (\text{as } s \to \pm\infty).$$

As $s \to +\infty$, the numerator tends to $a_2 - a_1$, but the denominator goes to $+\infty$. For $s \to -\infty$, the numerator becomes 0 and the denominator tends to μ. An analogous argument can be applied to the term $\tfrac{w'_\mu(s)-a_2}{w_\mu(s)-a_2 s}$. This yields the limits

$$\lim_{s\to-\infty} v'_\mu(s) = a_1 \text{ and } \lim_{s\to+\infty} v'_\mu(s) = a_2. \tag{3.10}$$

Consequently, since v'_μ is strictly increasing ($v''_\mu > 0$), we have

$$v'_\mu(\mathbb{R}) = (a_1, a_2) \subset (0, \infty). \tag{3.11}$$

As a result, v_μ itself is strictly increasing. Combining these facts along with the mean-value theorem yields (3.7). This completes the proof. ∎

We may now prove the following essential corollary.

Corollary 3.6. *For every $\mu > 0$, the log-barrier risk measure $\mathcal{R}_\mu$ is proper, closed, convex and satisfies properties* (2.2a)-(2.2d). *In addition, $\mathcal{R}_\mu : \mathcal{X} \to \mathbb{R}$, i.e. $\mathcal{R}_\mu$ is finite-valued on $L^1(\Omega, \mathcal{F}, \mathbb{P})$ and therefore, subdifferentiable.*

Proof. Let $X \in \mathcal{X}$ ($p = 1$). Then by (3.7) and the monotonicity of the expectation we have

$$\begin{aligned}\mathcal{R}_\mu(X) &= \inf_{t\in\mathbb{R}} \{t + \mathbb{E}[v_\mu(X - t)]\} \le \mathbb{E}[v_\mu(X) + v_\mu(0) - v_\mu(0)] \\ &\le \mathbb{E}[|v_\mu(X) - v_\mu(0)|] + |v_\mu(0)| \le a_2\mathbb{E}[|X|] + |v_\mu(0)| < +\infty.\end{aligned}$$

In order to apply Theorem 2.1, we recall from Lemma 3.3 that $v_\mu(s)$ can be replaced by

$$\widehat{v}_\mu(s) := v_\mu(s + d(\mu)) - d(\mu), \quad d(\mu) := \tfrac{2-a_1-a_2}{(1-a_1)(a_2-1)}\mu \in \mathbb{R}. \tag{3.12}$$

Clearly, $\widehat{v}_\mu(s)$ retains all the properties of v_μ that we proved in Proposition 3.5. It remains to show that $\widehat{v}_\mu$ fulfills (2.1). One readily shows by a simple calculation, cf. [23], that

$$\widehat{v}_\mu(0) = 0 \text{ and } \widehat{v}'_\mu(0) = 1. \tag{3.13}$$

Note that $\zeta(\mu)$ and the choice of $d(\mu)$ ensure that $\widehat{v}_\mu(0) = 0$. This and the strict convexity of $\widehat{v}_\mu$ imply

$$\widehat{v}_\mu(s) > s \text{ for all } s \in \mathbb{R} \setminus \{0\}.$$

The rest follows from Proposition 3.5 as an immediate consequence of Theorem 2.1. ∎

In order to obtain explicit optimality conditions suitable for the development of an optimization algorithm, we derive an explicit formula for $\partial\mathcal{R}_\mu$. We start by analyzing the log-barrier regret function. Afterwards, we prove that $\mathcal{R}_\mu$ is Hadamard differentiable.

Proposition 3.7. *Let $\mu > 0$ and define the log-barrier regret functional $\mathcal{V}_\mu : L^r(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ by*

$$\mathcal{V}_\mu(X) := \mathbb{E}[v_\mu(X)],$$

where $r \in [1, \infty]$. Then $\mathcal{V}_\mu$ is Hadamard differentiable. If $r > 1$, then $\mathcal{V}_\mu$ is continuously (Fréchet) differentiable. In both cases, the associated gradient takes the form

$$\nabla\mathcal{V}_\mu(X) = v'_\mu(X), \tag{3.14}$$

where $v'_\mu(X)$ is the superposition operator generated by the scalar function v'_μ.

Proof. The assertion follows a standard argument for differentiating integral functionals. We briefly sketch the main points here. Let $X, H \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ and $\tau > 0$. Then

$$\mathcal{V}_\mu(X + \tau H) - \mathcal{V}_\mu(X) = \mathbb{E}\left[v_\mu(X + \tau H) - v_\mu(X)\right].$$

Now, for $\mathbb{P}$-a.a. $\omega \in \Omega$, we have the pointwise limit

$$\tau^{-1}\left((v_\mu(X + \tau H))(\omega) - (v_\mu(X))(\omega)\right) \overset{\tau\downarrow 0}{=} (v'_\mu(X)H)(\omega).$$

In addition, it follows from (3.7) that

$$|\tau^{-1}\left((v_\mu(X + \tau H))(\omega) - (v_\mu(X))(\omega)\right)| \le a_2|H(\omega)| \in L^1(\Omega, \mathcal{F}, \mathbb{P}).$$

Hence, $\mathcal{V}_\mu$ is (Gâteaux) directionally differentiable. Furthermore, since for $\mathbb{P}$-a.a. $\omega \in \Omega$, we have

$$a_1 < v'_\mu(X(\omega)) < a_2, \tag{3.15}$$

$\mathcal{V}'_\mu(X; H)$ is continuous and linear in H and therefore, $\mathcal{V}_\mu$ is Gâteaux differentiable. Due to local Lipschitz continuity, $\mathcal{V}_\mu$ is Hadamard differentiable.

Finally, letting $r > 1$, we consider the superposition operator generated by v'_μ. By Proposition 3.5, v'_μ is a Carathéodory function. Then, using (3.15), the superposition operator generated by v'_μ maps all of $L^r(\Omega, \mathcal{F}, \mathbb{P})$ into $L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ and consequently $L^s(\Omega, \mathcal{F}, \mathbb{P})$, where $1/s + 1/r = 1$. The continuity of $\nabla \mathcal{V}_\mu(X) = v'_\mu(X)$ follows from Krasnoselskii's theorem, see, e.g., [2]. Therefore, $\mathcal{V}_\mu$ is continuously (Fréchet) differentiable, see, e.g., [9, pp. 35-36]. ◻

We can now obtain a more explicit formula for the gradient of the log-barrier risk measure.

Proposition 3.8. *Let $\mu > 0$ and $\mathcal{X} = L^1(\Omega, \mathcal{F}, \mathbb{P})$. Then $\mathcal{R}_\mu : \mathcal{X} \to \mathbb{R}$ is Hadamard differentiable with gradient*

$$\nabla \mathcal{R}_\mu(X) = v'_\mu(X - \mathcal{S}_\mu(X)) \in L^\infty(\Omega, \mathcal{F}, \mathbb{P}),$$

where $\mathcal{S}_\mu(X)$ is the associated statistic, i.e.,

$$\mathcal{S}_\mu(X) = \operatorname*{argmin}_{t \in \mathbb{R}} \{t + \mathbb{E}[v_\mu(X - t)]\}.$$

Proof. By Corollary 3.6, $\mathcal{R}_\mu$ is finite, closed, and convex and therefore, continuous, see, e.g., [19, Chap. 1. Thm. 2.5] or [46, Prop. 6.6]. It follows that $\partial \mathcal{R}_\mu(X) \neq \emptyset$ for all $X \in \mathcal{X}$. Next, fix $X \in \mathcal{X}$ and select an arbitrary $\vartheta \in \partial \mathcal{R}_\mu(X)$. By definition, we have

$$\mathcal{R}_\mu(Y) - \mathcal{R}_\mu(X) \geq \mathbb{E}[\vartheta(Y - X)], \quad \forall Y \in \mathcal{X}.$$

We can estimate the lefthand side of this inequality from above by using our knowledge of the statistic $\mathcal{S}_\mu(X)$. Consider the one-dimensional optimization problem

$$\inf_{t \in \mathbb{R}} \{\varphi_\mu(t) := \mathbb{E}[t + v_\mu(X - t)]\}.$$

By Proposition 3.5, φ_μ is strictly convex and differentiable. Therefore, since $\mathcal{S}_\mu(X)$ is a compact connected interval (cf. Theorem 2.1), it must be a singleton. It follows that

$$\begin{aligned} \mathbb{E}[\vartheta(Y - X)] &\leq \mathcal{S}_\mu(X) + \mathbb{E}[v_\mu(Y - \mathcal{S}_\mu(X))] - (\mathcal{S}_\mu(X) + \mathbb{E}[v_\mu(X - \mathcal{S}_\mu(X))]) \\ &= \mathbb{E}[v_\mu(Y - \mathcal{S}_\mu(X)) - v_\mu(X - \mathcal{S}_\mu(X))]. \end{aligned}$$

Setting $Y = X + \tau H$ for some $\tau > 0$ and $H \in \mathcal{X}$, we now have

$$\mathbb{E}[\vartheta H] \leq \tau^{-1} \mathbb{E}[v_\mu(X + \tau H - \mathcal{S}_\mu(X)) - v_\mu(X - \mathcal{S}_\mu(X))]$$

Passing to the limit as $\tau \downarrow 0$ yields

$$\mathbb{E}[\vartheta H] \leq \mathbb{E}[v'_\mu(X - \mathcal{S}_\mu(X)) H], \quad \forall H \in \mathcal{X}.$$

Since this holds for the entire space $\mathcal{X}$, we have $\vartheta = v'_\mu(X - \mathcal{S}_\mu(X))$. ◻

Recall that by the Fenchel-Young inequality, $\vartheta \in \partial \mathcal{R}_\mu(X)$ if and only if $\mathcal{R}_\mu(X) + \mathcal{R}_\mu^*(\vartheta) = \mathbb{E}[\vartheta X]$, where $\mathcal{R}_\mu^*$ is the usual Fenchel conjugate of $\mathcal{R}_\mu$. In particular, $\vartheta \in \operatorname{dom}(\mathcal{R}^*)$. One can then show, cf. [32, Prop B.4] that $\vartheta = v_\mu'(X - \mathcal{S}_\mu(X)) \in L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ fulfills

$$v_\mu'(X - \mathcal{S}_\mu(X)) \geq 0 \ \mathbb{P}\text{-a.e.}, \quad \mathbb{E}[v_\mu'(X - \mathcal{S}_\mu(X))] = 1, \quad \mathbb{E}[v_\mu^*(v_\mu'(X - \mathcal{S}_\mu(X)))] < +\infty.$$

This implies that ϑ (the so-called *risk indicator*) is a probability density.

4. Existence and Optimality Conditions.

In this section, we use the analysis of $\mathcal{R}_\mu$ from the previous section to prove the existence of minimizers and derive explicit optimality conditions.

4.1. Random Fields and Objective Functionals.

As noted in the introduction, $u = S(z)$ is the random field solution for some PDE or system of PDEs with random inputs. It is essential that S, as a mapping from z into some Bochner space, fulfills sufficient continuity and differentiability properties in order to guarantee existence of solutions to (1.1). As in [32], we make the following assumptions:

ASSUMPTION 4.1 (Properties of the solution map).

1. *$S(z)$ is unique for all $z \in \mathcal{Z}_{\mathrm{ad}}$.*
2. *$S(z) : \Omega \to U$ is strongly $\mathcal{F}$-measurable for all $z \in \mathcal{Z}_{\mathrm{ad}}$.*
3. *There exist a nonnegative increasing function $\rho : [0, \infty) \to [0, \infty)$ and a nonnegative random variable $C \in L^q(\Omega, \mathcal{F}, \mathbb{P})$ with $q \in [1, \infty]$ satisfying*

$$\|S(z)\|_U \leq C\rho(\|z\|_Z) \quad \mathbb{P}\text{-}a.e. \quad \forall\, z \in \mathcal{Z}_{\mathrm{ad}}.$$

4. *If $z_n \rightharpoonup z$ in $\mathcal{Z}_{\mathrm{ad}}$, then $S(z_n) \rightharpoonup S(z)$ in U $\mathbb{P}$-a.e.*
5. *There exists an open set $V \subseteq Z$ with $\mathcal{Z}_{\mathrm{ad}} \subseteq V$ such that the solution map $V \ni z \mapsto S(z) : V \to L^q(\Omega, \mathcal{F}, \mathbb{P}; U)$ is continuously differentiable.*

For readability, we will denote this "stochastic" state space by

$$\mathcal{U} := L^q(\Omega, \mathcal{F}, \mathbb{P}; U),$$

where q is from Assumption 4.1.3. Moreover, we note that the first three conditions ensure $S(z) \in \mathcal{U}$ for any decision $z \in \mathcal{Z}_{\mathrm{ad}}$. In fact, Assumptions 4.1.1.-4. imply that S is weakly continuous in sense that

$$z_n \overset{Z}{\rightharpoonup} z \Longrightarrow S(z_n) \overset{\mathcal{U}}{\rightharpoonup} S(z).$$

These conditions will generally be enough to prove existence of minimizers (As. 4.1.1.-4.) for (1.1) and derive optimality conditions (As. 4.1.5.) as discussed in [32]. Furthermore, they can be readily verified for a wide variety of random PDEs, e.g., linear elliptic PDE with random coefficients. The situation is potentially more involved for nonlinear PDE, see e.g., [33] for a recent study.

Turning now to the objective function, we will assume that $\mathcal{J}$ is generated by a parametrized function $J : U \times \Omega \to \mathbb{R}$. Recall that for some mapping $u : \Omega \to U$, J generates a nonlinear superposition operator

$$[\mathcal{J}(u)](\omega) := J(u(\omega), \omega),$$

In order to prove existence of a solution and derive optimality conditions, we will need the following assumptions.

Assumption 4.2 (Properties of $J : U \times \Omega \to \mathbb{R}$).

1. *J is a Carathéodory function, i.e., $J(\cdot,\omega)$ is continuous for $\mathbb{P}$-a.a. $\omega \in \Omega$ and $J(u,\cdot)$ is measurable for all $u \in U$.*
2. *If $1 \le p, q < \infty$, then there exists $a \in L^p(\Omega,\mathcal{F},\mathbb{P})$ with $a \ge 0$ $\mathbb{P}$-a.e. and $c > 0$ such that*

$$|J(u,\omega)| \le a(\omega) + c\|u\|_U^{q/p} \quad \forall\, u \in U. \tag{4.1}$$

If $1 \le p < \infty$ and $q = \infty$, then the uniform boundedness condition holds: for all $c > 0$ there exists $\gamma = \gamma(c) \in L^p(\Omega,\mathcal{F},\mathbb{P})$ such that

$$|J(u,\omega)| \le \gamma(\omega) \ \text{ for } \mathbb{P}\text{-a.a. } \omega \in \Omega \quad \forall\, u \in U,\ \|u\|_U \le c. \tag{4.2}$$

3. *$J(\cdot,\omega)$ is convex for $\mathbb{P}$-a.a. $\omega \in \Omega$.*

Note that Assumptions 4.2.1.-2. guarantee by Krasnoselskii's theorem that $\mathcal{J} : \mathcal{U} \to \mathcal{X}$ is continuous under appropriate assumptions on p, q, see e.g., [2]. Together with Assumption 4.2.3., we can prove that $\mathcal{J} : \mathcal{U} \to \mathcal{X}$ is Gâteaux directionally differentiable, [32, Thm. 3.9]. For further smoothness, we require at least local Lipschitz continuity of $\mathcal{J} : \mathcal{U} \to \mathcal{X}$ and more structure.

4.2. Existence and Optimality Theory. In this subsection, we briefly state the necessary general existence and optimality results.

Theorem 4.3. *Let Assumptions 4.1 and 4.2 hold. Moreover, suppose that $\mathcal{R} : \mathcal{X} \to \overline{\mathbb{R}}$ is generated as in Theorem 2.1 and $\wp : Z \to \overline{\mathbb{R}}$ be proper, closed, and convex. Finally, suppose that either $\mathcal{Z}_{\text{ad}}$ is bounded or $z \mapsto \mathcal{R}(\mathcal{J}(S(z))) + \wp(z)$ is radially unbounded on $\mathcal{Z}_{\text{ad}}$, i.e., $z_k \in \mathcal{Z}_{\text{ad}}$ such that $\|z_k\|_Z \to +\infty$ implies $\mathcal{R}(\mathcal{J}(S(z_k))) + \wp(z_k) \to +\infty$.*

1. *If either $\mathcal{R}$ is finite-valued on all of $\mathcal{X}$ or $\operatorname{int}\operatorname{dom}(\mathcal{R}) \neq \emptyset$, then (1.1) has an optimal solution $z^\star$.*
2. *If, in addition, $\mathcal{J} : \mathcal{U} \to \mathcal{X}$ is locally Lipschitz and $\wp$ is Gâteaux directionally differentiable, then there exists $\vartheta^\star \in \partial\mathcal{R}(\mathcal{J}(S(z^\star)))$ such that following first-order optimality condition holds:*

$$\mathbb{E}[\mathcal{J}'(S(z^\star); S'(z^\star)(z - z^\star))\,\vartheta^\star] + \wp'(z^\star; z - z^\star) \ge 0 \quad \forall\, z \in \mathcal{Z}_{\text{ad}}. \tag{4.3}$$

where $\partial\mathcal{R}(\mathcal{J}(S(z^\star))) \subset \mathcal{X}^*$ is the usual subdifferential of $\mathcal{R}$ at $\mathcal{J}(S(z))$.

Proof. See [32, Prop. 3.12, Prop. 3.13]. ∎

4.3. Analysis of the Log-Barrier Optimization Problems.

Theorem 4.4. *Let Assumptions 4.1 and 4.2 hold and set $\mathcal{X} = L^1(\Omega,\mathcal{F},\mathbb{P})$. Furthermore, suppose that $\wp : Z \to \overline{\mathbb{R}}$ is be proper, closed, and convex. Then for every $\mu > 0$, the optimization problem*

$$\min_{z \in \mathcal{Z}_{\text{ad}}} \mathcal{R}_\mu(\mathcal{J}(S(z))) + \wp(z) \tag{4.4}$$

admits a solution $z^\star \in \mathcal{Z}_{\text{ad}}$ provided either Z_{ad} is bounded or $z \mapsto \mathcal{R}_\mu(\mathcal{J}(S(z))) + \wp(z)$ is radially unbounded on $\mathcal{Z}_{\text{ad}}$.

Proof. In light of Corollary 3.6, this is a direct consequence of Theorem 4.3. ∎

Similarly, we can also leverage the results of Section 3 to derive optimality conditions.

THEOREM 4.5. *Let Assumptions 4.1 and 4.2 hold, set $\mathcal{X} = L^1(\Omega, \mathcal{F}, \mathbb{P})$, and fix $\mu > 0$. Furthermore, suppose that $\wp : Z \to \overline{\mathbb{R}}$ is proper, closed, and convex and assume an optimal solution $z^\star$ to (4.4) exists. If $\mathcal{J} : \mathcal{U} \to \mathcal{X}$ is locally Lipschitz and $\wp$ is Gâteaux directionally differentiable, then there exists $t^\star \in \mathbb{R}$ such that*

$$\mathbb{E}[\mathcal{J}'(S(z^\star); S'(z^\star)(z - z^\star))\, v'_\mu(\mathcal{J}(S(z^\star)) - t^\star)] + \wp'(z^\star; z - z^\star) \geq 0 \quad \forall\, z \in \mathcal{Z}_{\text{ad}}, \tag{4.5a}$$

$$\mathbb{E}[v'_\mu(\mathcal{J}(S(z^\star)) - t^\star)] = 1. \tag{4.5b}$$

Proof. According to Theorem 4.3, there exists $\vartheta^\star \in \partial\mathcal{R}(\mathcal{J}(S(z^\star)))$ such that

$$\mathbb{E}[\mathcal{J}'(S(z^\star); S'(z^\star)(z - z^\star))\, \vartheta^\star] + \wp'(z^\star; z - z^\star) \geq 0,\ \forall\, z \in \mathcal{Z}_{\text{ad}}.$$

Moreover, by Proposition 3.8,

$$\vartheta^\star = v'_\mu(\mathcal{J}(S(z^\star)) - \mathcal{S}_\mu(\mathcal{J}(S(z^\star))))$$

and

$$\mathcal{S}_\mu(\mathcal{J}(S(z^\star)) = \underset{t \in \mathbb{R}}{\operatorname{argmin}}\, \{t + \mathbb{E}[v_\mu(\mathcal{J}(S(z^\star)) - t)]\}.$$

The (unique) statistic $\mathcal{S}_\mu(\mathcal{J}(S(z^\star))$ can be equivalently described by the first-order necessary and sufficient conditions for this one-dimensional optimization problem, which are given by (4.5b). The rest follows by substitution with $t^\star = \mathcal{S}_\mu(\mathcal{J}(S(z^\star))$. ∎

The conditions (4.5) are rather abstract. In order to develop a viable, i.e., implementable, numerical optimization algorithm, we need to "unfold" these conditions to obtain a more amenable optimality system. In the sequel, we assume, in addition to the hypotheses of Theorem 4.5, that $\mathcal{J}$ and $\wp$ admit gradients $\nabla\mathcal{J}$ and $\nabla\wp$, respectively, and that Z is a real Hilbert space. Furthermore, we set

$$g(z) := (\mathcal{J} \circ S)(z) \text{ and } g'(z) = \mathcal{J}'(S(z)) \circ S'(z).$$

Then by substitution into (4.5), we have

$$\langle \mathbb{E}[v'_\mu(g(z^\star) - t^\star) g'(z^\star)] + \wp'(z^\star), z - z^\star \rangle \geq 0 \quad \forall\, z \in \mathcal{Z}_{\text{ad}}.$$

Since Z is assumed to be a Hilbert space and $\mathcal{Z}_{\text{ad}}$ is a nonempty, closed, and convex set, the previous variational inequality can be rewritten as

$$z^\star = \operatorname{Proj}_{\mathcal{Z}_{\text{ad}}}\left(z^\star - c(\mathbb{E}[v'_\mu(g(z^\star) - t^\star)\nabla g(z^\star)] + \nabla\wp(z^\star)\right), \quad c > 0, \tag{4.6}$$

where $\nabla g(z^\star) = S'(z^\star)^*\nabla\mathcal{J}(S(z^\star))$ and $\nabla\wp(z^\star)$ are the Riesz representations of the derivatives $g'(z^\star)$ and $\wp'(z^\star)$ in Z. Continuing, if $\mathcal{Z}_{\text{ad}} = Z$, then we obtain the usual gradient equation (in Z^*):

$$\mathbb{E}[v'_\mu(g(z^\star) - t^\star)\nabla g(z^\star)] + \nabla\wp(z^\star) = 0. \tag{4.7}$$

Next, we recall that

$$W_{g(z^\star),t^\star} = \mu + \tfrac{1}{2}\left((a_1 + a_2)(g(z^\star) - t^\star) + \sqrt{(a_2 - a_1)^2(g(z^\star) - t^\star)^2 + 4\mu^2}\right) \tag{4.8}$$

is the (pointwise a.e. unique) solution to

$$1 - \tfrac{\mu}{w - a_1(g(z^\star) - t^\star)} - \tfrac{\mu}{w - a_2(g(z^\star) - t^\star)} = 0,$$

see (3.6). Thus, rather than using the explicit form for $W_{g(z^\star),t^\star}$ we replace it by (re)introducing the variable $W \in \mathcal{X}$ and adding the equation

$$1 - \tfrac{\mu}{W^\star - a_1(g(z^\star)-t^\star)} - \tfrac{\mu}{W^\star - a_2(g(z^\star)-t^\star)} = 0 \quad \mathbb{P}\text{-a.e.} \tag{4.9}$$

to the optimality system. Note that one can derive the following pointwise inequality:

$$W^\star - a_i(g(z^\star) - t^\star) > \mu. \tag{4.10}$$

We can now simplify the risk indicator formula. Indeed, given $W^\star$, we have

$$\begin{aligned}
v'_\mu(g(z^\star) - t^\star) &= w'_\mu(g(z^\star) - t^\star) - \mu \tfrac{w'_\mu(g(z^\star)-t^\star)-a_1}{W^\star - a_1(g(z^\star)-t^\star)} - \mu \tfrac{w'_\mu(g(z^\star)-t^\star)-a_2}{W^\star - a_2(g(z^\star)-t^\star)} \\
&= w'_\mu(g(z^\star) - t^\star) \left(1 - \tfrac{\mu}{W^\star - a_1(g(z^\star)-t^\star)} - \tfrac{\mu}{W^\star - a_2(g(z^\star)-t^\star)}\right) \\
&\qquad + \mu \left(\tfrac{a_1}{W^\star - a_1(g(z^\star)-t^\star)} + \tfrac{a_2}{W^\star - a_2(g(z^\star)-t^\star)}\right) \\
&= \mu \left(\tfrac{a_1}{W^\star - a_1(g(z^\star)-t^\star)} + \tfrac{a_2}{W^\star - a_2(g(z^\star)-t^\star)}\right).
\end{aligned}$$

This leads to the following result.

PROPOSITION 4.6. *In addition to the hypotheses of Theorem 4.5, assume that Z is a real Hilbert space and $\mathcal{J}$ and $\wp$ admit gradients $\nabla\mathcal{J}$ and $\nabla\wp$, respectively. Then there exist $t^\star \in \mathbb{R}$, $W^\star \in \mathcal{X}$ such that*

$$z^\star - \mathrm{Proj}_{\mathcal{Z}_{\mathrm{ad}}}\Big(z^\star - c\Big(\mathbb{E}\Big[\mu \sum_{i=1,2} \tfrac{a_i \nabla g(z^\star)}{W^\star - a_i(g(z^\star)-t^\star)}\Big] + \nabla\wp(z^\star)\Big)\Big) = 0, \quad c > 0, \tag{4.11a}$$

$$\mu \sum_{i=1,2} \tfrac{1}{W^\star - a_i(g(z^\star)-t^\star)} - 1 = 0 \quad \mathbb{P}\text{-}a.e., \tag{4.11b}$$

$$\mu\, \mathbb{E}\Big[\sum_{i=1,2} \tfrac{a_i}{W^\star - a_i(g(z^\star)-t^\star)}\Big] - 1 = 0 \tag{4.11c}$$

Proof. This is a direct result of the preceeding discussion. ∎

As an alternative to (4.11), we can also consider the equivalent primal dual optimality system by introducing slack variables ν_i given by $\nu_i := \frac{\mu}{W^\star - a_i(g(z^\star)-t^\star)}$. In light of (4.10), we have $\nu_i \in L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ and $\nu_i > 0$ $\mathbb{P}$-a.e. as would perhaps be expected since $\vartheta^\star \in \mathcal{X}^* = L^\infty(\Omega, \mathcal{F}, \mathbb{P})$.

PROPOSITION 4.7. *In addition to the hypotheses of Theorem 4.5, assume that Z is a real Hilbert space and that $\mathcal{J}$ and $\wp$ admit gradients $\nabla\mathcal{J}$ and $\nabla\wp$, respectively. Then there exist $t^\star \in \mathbb{R}$, $W^\star \in \mathcal{X}$, $\nu_i^\star \in \mathcal{X}^*$ $(i = 1, 2)$ such that*

$$z^\star - \mathrm{Proj}_{\mathcal{Z}_{\mathrm{ad}}}\left(z^\star - c\left(\mathbb{E}\left[(a_1\nu_1^\star + a_2\nu_2^\star)\nabla g(z^\star)\right] + \nabla\wp(z^\star)\right)\right) = 0, \quad c > 0, \tag{4.12a}$$

$$(\nu_1^\star + \nu_2^\star) - 1 = 0 \quad \mathbb{P}\text{-}a.e., \tag{4.12b}$$

$$\mathbb{E}\left[(a_1\nu_1^\star + a_2\nu_2^\star)\right] - 1 = 0, \tag{4.12c}$$

$$\nu_i^\star(W^\star - a_i(g(z^\star) - t^\star)) = \mu \quad \mathbb{P}\text{-}a.e.. \tag{4.12d}$$

5. Newton System.

Let $\mathcal{Z}_{\mathrm{ad}} = Z$ and set $\mathcal{X} = L^\infty(\Omega, \mathcal{F}, \mathbb{P})$. Assume that both $g : Z \to \mathcal{X}$ and $\wp : Z \to \mathbb{R}$ are twice continuously differentiable. (4.12) with $c = 1$

reads

$$
\begin{aligned}
&(5.1\text{a}) \qquad F^1(z,t,W,\nu_1,\nu_2) := \mathbb{E}\left[(a_1\nu_1 + a_2\nu_2)\nabla g(z)\right] + \nabla\wp(z) = 0 \in Z,\\
&(5.1\text{b}) \qquad F^2(z,t,W,\nu_1,\nu_2) := \mathbb{E}\left[(a_1\nu_1 + a_2\nu_2)\right] - 1 = 0 \in \mathbb{R},\\
&(5.1\text{c}) \qquad F^3(z,t,W,\nu_1,\nu_2) := (\nu_1 + \nu_2) - 1 = 0 \in \mathcal{X},\\
&(5.1\text{d}) \qquad F^4(z,t,W,\nu_1,\nu_2) := \nu_1(W - a_1(g(z) - t)) - \mu = 0 \in \mathcal{X},\\
&(5.1\text{e}) \qquad F^5(z,t,W,\nu_1,\nu_2) := \nu_2(W - a_2(g(z) - t)) - \mu = 0 \in \mathcal{X}.
\end{aligned}
$$

From the above considerations we have $W^\star \in \mathcal{X}$ and $W^\star - a_i(g(z^\star) - t^\star) \geq \mu$ a.s. as well as $\nu_i^\star \in \mathcal{X}$ and $0 < \frac{\mu}{\|W^\star - a_i(g(z^\star)-t^\star)\|_{\mathcal{X}}} \leq \nu_i^\star \leq 1$ a.s. due to (5.1d) and (5.1e) for the solution $(z^\star, t^\star, W^\star, \nu_1^\star, \nu_2^\star)$ of this system. Therefore, it makes sense to keep $V_i := W - a_i(g(z) - t)$ and ν_i uniformly positive during the solution process.

Lemma 5.1. *The function $F : Z \times \mathbb{R} \times \mathcal{X} \times \mathcal{X} \times \mathcal{X} \to Z \times \mathbb{R} \times \mathcal{X} \times \mathcal{X} \times \mathcal{X}$ defined in (5.1) is continuously differentiable. Leaving off (z, t, W, ν_1, ν_2), we have*

$$F^1_z s = \mathbb{E}\left[(a_1\nu_1 + a_2\nu_2)\nabla^2 g(z)s\right] + \nabla^2\wp(z)s \in Z,$$

$$
\begin{array}{rclrcl}
F^1_{\nu_1}\delta_1 &=& \mathbb{E}[a_1\delta_1\nabla g(z)] \in Z, & F^1_{\nu_2}\delta_2 &=& \mathbb{E}[a_2\delta_2\nabla g(z)] \in Z,\\
F^2_{\nu_1}\delta_1 &=& \mathbb{E}[a_1\delta_1] \in \mathbb{R}, & F^2_{\nu_2}\delta_2 &=& \mathbb{E}[a_2\delta_2] \in \mathbb{R},\\
F^3_{\nu_1}\delta_1 &=& \delta_1 \in \mathcal{X}, & F^3_{\nu_2}\delta_2 &=& \delta_2 \in \mathcal{X},
\end{array}
$$

$$
\begin{array}{rclrcl}
F^4_z s &=& -a_1(\nabla g(z), s)_Z\nu_1 \in \mathcal{X}, & F^5_z s &=& -a_2(\nabla g(z), s)_Z\nu_2 \in \mathcal{X},\\
F^4_t\tau &=& a_1\tau\nu_1 \in \mathcal{X}, & F^5_t\tau &=& a_2\tau\nu_2 \in \mathcal{X},\\
F^4_t S &=& \nu_1 S \in \mathcal{X}, & F^5_t S &=& \nu_2 S \in \mathcal{X},\\
F^4_{\nu_1}\delta_1 &=& (W - a_1(g(z) - t))\delta_1 \in \mathcal{X}, & F^5_{\nu_2}\delta_2 &=& (W - a_2(g(z) - t))\delta_2 \in \mathcal{X},
\end{array}
$$

and the remaining derivatives are zero.

Proof. Note that the pointwise multiplication operator $\mathcal{X} \times \mathcal{X} \ni (V, W) \mapsto VW \in \mathcal{X}$ is continuously differentiable and its derivative w.r.t. V is represented by W and vice versa. Applying the chain rule yields the desired result. ∎

We now write $\nabla^2 h(z, \nu_1, \nu_2) := \mathbb{E}\left[(a_1\nu_1 + a_2\nu_2)\nabla^2 g(z)\right] + \nabla^2\wp(z)$. With the computed derivatives, the Newton equation reads

$$
\begin{pmatrix}
\nabla^2 h(z,\nu_1,\nu_2) & 0 & 0 & \mathbb{E}[a_1\nabla g(z)\cdot] & \mathbb{E}[a_2\nabla g(z)\cdot]\\
0 & 0 & 0 & \mathbb{E}[a_1\cdot] & \mathbb{E}[a_2\cdot]\\
0 & 0 & 0 & I & I\\
-a_1\nu_1(\nabla g(z),\cdot)_Z & a_1\nu_1 & \nu_1 & V_1 & 0\\
-a_2\nu_2(\nabla g(z),\cdot)_Z & a_2\nu_2 & \nu_2 & 0 & V_2
\end{pmatrix}
\begin{pmatrix} s\\ \tau\\ S\\ \delta_1\\ \delta_2 \end{pmatrix} = -F
$$

As long as ν_i is uniformly positive a.s., i.e., $\nu_i^{-1} \in \mathcal{X}$, we can multiply the fourth line by $-\nu_1^{-1}$ pointwisely, the fifth line by $-\nu_2^{-1}$, and multiply the second and third one by -1 to obtain the equivalent symmetric system

(5.2)

$$
\begin{pmatrix}
\nabla^2 h(z,\nu_1,\nu_2) & 0 & 0 & \mathbb{E}[a_1\nabla g(z)\cdot] & \mathbb{E}[a_2\nabla g(z)\cdot]\\
0 & 0 & 0 & -\mathbb{E}[a_1\cdot] & -\mathbb{E}[a_2\cdot]\\
0 & 0 & 0 & -I & -I\\
a_1(\nabla g(z),\cdot)_Z & -a_1 & -I & -\nu_1^{-1}V_1 & 0\\
a_2(\nabla g(z),\cdot)_Z & -a_2 & -I & 0 & -\nu_2^{-1}V_2
\end{pmatrix}
\begin{pmatrix} s\\ \tau\\ S\\ \delta_1\\ \delta_2 \end{pmatrix}
=
\begin{pmatrix}
-\mathbb{E}[(a_1\nu_1+a_2\nu_2)\nabla g(z)] - \nabla\wp(z)\\
\mathbb{E}[a_1\nu_1+a_2\nu_2] - 1\\
(\nu_1+\nu_2) - 1\\
(W - a_1(g(z)-t)) - \mu\nu_1^{-1}\\
(W - a_2(g(z)-t)) - \mu\nu_2^{-1}
\end{pmatrix}
$$

Lemma 5.2. *Let $\nu_i \in \mathcal{X}$ and $V_i = W - a_i(g(z) - t) \in \mathcal{X}$ be uniformly positive for $i \in \{1, 2\}$. If the operator*

$$\nabla^2 h(z, \nu_1, \nu_2) = \mathbb{E}\left[(a_1\nu_1 + a_2\nu_2)\nabla^2 g(z)\right] + \nabla^2 \wp(z) : Z \to Z$$

is coercive, the Newton operator defined in (5.2) has a bounded inverse.

Proof. We apply the bounded inverse theorem. Since $\nu_i^{-1} V_i \in \mathcal{X}$, the operator is linear and bounded as a map from $Z \times \mathbb{R} \times \mathcal{X} \times \mathcal{X} \times \mathcal{X}$ to itself. Therefore, it is sufficient to show that it is bijective. Consider the Newton equation with general right-hand side $(r_z, r_t, r_W, r_{\nu_1}, r_{\nu_2}) \in Z \times \mathbb{R} \times \mathcal{X} \times \mathcal{X} \times \mathcal{X}$. Since $T_i := \nu_i V_i^{-1} \in \mathcal{X}$ by assumption, the last two lines can be uniquely solved for $\delta_i = -T_i r_{\nu_i} + a_i T_i (\nabla g(z), s)_Z - a_i \tau T_i - T_i S$, respectively, given (s, τ, S). This yields the reduced, symmetric Newton system

$$\begin{aligned}
&\begin{pmatrix} \star & -\mathbb{E}[(a_1^2 T_1 + a_2^2 T_2)\nabla g(z)] & -\mathbb{E}[(a_1 T_1 + a_2 T_2)\nabla g(z)\cdot] \\ -\mathbb{E}[(a_1^2 T_1 + a_2^2 T_2)(\nabla g(z), \cdot)_Z] & (a_1^2 \mathbb{E}[T_1] + a_2^2 \mathbb{E}[T_2]) & \mathbb{E}[(a_1 T_1 + a_2 T_2)\cdot] \\ -(a_1 T_1 + a_2 T_2)(\nabla g(z), \cdot)_Z & a_1 T_1 + a_2 T_2 & T_1 + T_2 \end{pmatrix} \begin{pmatrix} s \\ \tau \\ S \end{pmatrix} \\
&= \begin{pmatrix} r_z + \mathbb{E}[(a_1 T_1 r_{\nu_1} + a_2 T_2 r_{\nu_2})\nabla g(z)] \\ r_t - \mathbb{E}[a_1 T_1 r_{\nu_1}] - \mathbb{E}[a_2 T_2 r_{\nu_2}] \\ r_W - T_1 r_{\nu_1} - T_2 r_{\nu_2} \end{pmatrix} =: \begin{pmatrix} \hat{r}_z \\ \hat{r}_t \\ \hat{r}_W \end{pmatrix}
\end{aligned} \tag{5.3}$$

with $\star = \nabla^2 h(z, \nu_1, \nu_2) + \mathbb{E}[(a_1^2 T_1 + a_2^2 T_2)(\nabla g(z), \cdot)_Z \nabla g(z)]$. T_i are uniformly positive. Hence, the third row can be solved for

$$S = (T_1 + T_2)^{-1}\big(r_W - T_1 r_{\nu_1} - T_2 r_{\nu_2} + (a_1 T_1 + a_2 T_2)(\nabla g(z), s)_Z - a_1 \tau T_1 - a_2 \tau T_2\big).$$

This yields the further reduced, symmetric system

$$\begin{pmatrix} *_{11} & *_{12} \\ *_{21} & *_{22} \end{pmatrix} \begin{pmatrix} s \\ \tau \end{pmatrix} = \begin{pmatrix} \hat{r}_z + \mathbb{E}[(a_1 T_1 + a_2 T_2)(T_1 + T_2)^{-1}(r_W - T_1 r_{\nu_1} - T_2 r_{\nu_2})\nabla g(z)] \\ \hat{r}_t - \mathbb{E}[(a_1 T_1 + a_2 T_2)(T_1 + T_2)^{-1}(r_W - T_1 r_{\nu_1} - T_2 r_{\nu_2})] \end{pmatrix} =: \begin{pmatrix} \tilde{r}_z \\ \tilde{r}_t \end{pmatrix} \tag{5.4}$$

with

$$\begin{aligned}
*_{11} &= \nabla^2 h(z, \nu_1, \nu_2) + \mathbb{E}[(a_1^2 T_1 + a_2^2 T_2)(\nabla g(z), \cdot)_Z \nabla g(z)] \\
&\quad - \mathbb{E}[(a_1 T_1 + a_2 T_2)^2 (T_1 + T_2)^{-1} (\nabla g(z), \cdot)_Z \nabla g(z)] \\
&= \nabla^2 h(z, \nu_1, \nu_2) + (a_2 - a_1)^2 \mathbb{E}[U(\nabla g(z), \cdot)_Z \nabla g(z)], \\
*_{12} &= -\mathbb{E}[(a_1^2 T_1 + a_2^2 T_2)\nabla g(z)] + \mathbb{E}[(a_1 T_1 + a_2 T_2)^2 (T_1 + T_2)^{-1} \nabla g(z)] \\
&= -(a_2 - a_1)^2 \mathbb{E}[U \nabla g(z)], \\
*_{21} &= -(a_2 - a_1)^2 \mathbb{E}[U(\nabla g(z), \cdot)_Z] \\
*_{22} &= (a_1^2 \mathbb{E}[T_1] + a_2^2 \mathbb{E}[T_2]) - \mathbb{E}[(a_1 T_1 + a_2 T_2)^2 (T_1 + T_2)^{-1}] = (a_2 - a_1)^2 \mathbb{E}[U],
\end{aligned}$$

where $U := T_1 T_2 (T_1 + T_2)^{-1} = (T_1^{-1} + T_2^{-1})^{-1} = \big(\nu_1^{-1} V_1 + \nu_2^{-1} V_2\big)^{-1}$. This function is uniformly positive by assumption and therefore $*_{22} > 0$ so that the system can be solved for

$$\tau = (a_2 - a_1)^{-2} \mathbb{E}[U]^{-1} \big(\tilde{r}_t + (a_2 - a_1)^2 \mathbb{E}[U(\nabla g(z), s)_Z]\big).$$

This gives the equation

$$\begin{aligned}
&\nabla^2 h(z, \nu_1, \nu_2) s + (a_2 - a_1)^2 \mathbb{E}[U(\nabla g(z), s)_Z \nabla g(z)] \\
&\qquad - (a_2 - a_1)^2 \mathbb{E}[U]^{-1}\, \mathbb{E}[U(\nabla g(z), s)_Z]\, \mathbb{E}[U \nabla g(z)] \\
&= \tilde{r}_z + \mathbb{E}[U]^{-1} \mathbb{E}[U \nabla g(z)] \big(\tilde{r}_t + (a_2 - a_1)^2 \mathbb{E}[U(\nabla g(z), s)_Z]\big)
\end{aligned} \tag{5.5}$$

for the control step s. Let now $\mathbb{E}_U[X] := \mathbb{E}[U]^{-1}\mathbb{E}[UX]$ be the expectation w.r.t. the probability measure induced by the random variable $\mathbb{E}[U]^{-1}U$. With this definition, the left-hand side operator applied to s is

$$\begin{aligned}&\nabla^2 h(z,\nu_1,\nu_2)s + (a_2-a_1)^2\mathbb{E}[U]\Big(\mathbb{E}_U[(\nabla g(z),s)_Z\nabla g(z)] - \mathbb{E}_U[(\nabla g(z),s)_Z]\,\mathbb{E}_U[\nabla g(z)]\Big)\\ &= \nabla^2 h(z,\nu_1,\nu_2)s + (a_2-a_1)^2\mathbb{E}[U]\,\mathrm{Cov}_U[(\nabla g(z),s)_Z,\nabla g(z)].\end{aligned}$$

Taking the Z inner product of this quantity and s yields

$$(\nabla^2 h(z,\nu_1,\nu_2)s,s)_Z + (a_2-a_1)^2\mathbb{E}[U]\,\mathrm{Var}_U[(\nabla g(z),s)_Z] \ge (\nabla^2 h(z,\nu_1,\nu_2)s,s)_Z \ge \gamma\|s\|_Z^2$$

for some $\gamma > 0$ by assumption. Hence, (5.5) has a unique solution s, from which we can compute the unique solution $(s,\tau,S,\delta_1,\delta_2)$ of the full Newton system from the above considerations. ◻

REMARK 5.3. *If ν_1 and ν_2 solve* (5.1d) *and* (5.1e)*, respectively, we have $T_i = \mu(W - a_i(g(z)-z))^{-2}$. Inserting this into the reduced Newton system* (5.3) *yields exactly the barrier-Newton system for the reduced version of* (5.1)*, i.e., the one where* (5.1d) *are* (5.1e) *solved for ν_i and the result is inserted into the remaining equations. Analogously, we can additionally solve* (5.1c) *for W using* (4.8) *and reduce the system further. The Newton equation for this system is then of the form* (5.4)*.*

REMARK 5.4. *The assumptions in Lemma 5.2 are very natural, at least for convex optimal control problems: The uniform positivity of the variables is ensured during the algorithm. If $\nabla^2 g(z)$ is positive (semidefinite) a.s. (e.g., if g is the convex reduced tracking term) and $\nabla^2\wp(z)$ is coercive (e.g., if $\wp(z) = \frac{\alpha}{2}\|z\|_Z^2$), the operator is coercive.*

6. Γ-Convergence of $\mathcal{R}_\mu$ to $\mathcal{R}$.

In order to argue that solutions of the approximating optimization problems converge to a solution of the original risk-averse optimization problem, we make use of several techniques from the theory of Γ-convergence, see, e.g., [15]. We recall that a sequence of functionals $\{\varphi_k\}$ on a topological space $\mathcal{X}$ Γ-converges to a functional $\varphi : \mathcal{X} \to \overline{\mathbb{R}}$, denoted by $\varphi_k \xrightarrow{\Gamma} \varphi$ provided the following two conditions hold:

1. $\forall\, x \in \mathcal{X}$, $\forall\, \{x_k\} \subset \mathcal{X}$ such that $x_k \to x$ we have $\liminf_k \varphi_k(x_k) \ge \varphi(x)$.
2. $\forall\, x \in \mathcal{X}$, $\exists\,\{x_k\}$ such that $x_k \to x$ and $\limsup_k \varphi_k(x_k) \le \varphi(x)$.

Note the theory is sufficiently general so that we may use rather coarse topologies on the spaces of random variables if necessary. We make the standing assumptions throughout that $\mathcal{X}$ is a topological vector space with the property that

$$\mathcal{X} \subset L^1(\Omega,\mathcal{F},\mathbb{P}).$$

In order to prove that $\mathcal{R}_\mu \xrightarrow{\Gamma} \mathcal{R}$ we use a result that combines several statements and remarks from [15]. For convenience, we state this here:

PROPOSITION 6.1. *Let $\mathcal{X}$ be a topological space and suppose that $\{F_k\}$ with $F_k : \mathcal{X} \to \overline{\mathbb{R}}$ is a sequence of lower-semicontinuous functionals. If $\{F_k\}$ is an increasing sequence of functionals that converges pointwise to F, then F is lower-semicontinuous and $F_k \xrightarrow{\Gamma} F$.*

Proof. This follows from [15, Prop. 5.4] as pointed out in [15, Remark 5.5]. ◻

We will need the following technical lemma concerning the smoothed scalar regret functions. As argued above, we use the shifted smoothed scalar regret function $\widehat{v}_\mu$ to generate $\mathcal{R}_\mu$.

LEMMA 6.2. *Let $\mu > 0$ and $\widehat{v}_\mu : \mathbb{R} \to \mathbb{R}$ be defined as in (3.12), (3.5) by $\widehat{v}_\mu(s) := v_\mu(s + d(\mu)) - d(\mu)$ with $d(\mu) = \frac{2-a_1-a_2}{(1-a_1)(a_2-1)}\mu$. Then the following properties hold:*

1. *$\widehat{v}_\mu(s) \le v(s)$ for all $s \in \mathbb{R}$.*
2. *$\lim_{\mu\to 0^+} \widehat{v}_\mu(s) = v(s)$ for all $s \in \mathbb{R}$.*
3. *$|\widehat{v}_\mu(s) - \widehat{v}_\mu(s')| \le a_2\, |s - s'|$ for all $s, s' \in \mathbb{R}$.*
4. *For all $\mu, \nu > 0$ such that $\mu \le \nu$ we have $\widehat{v}_\nu(s) \le \widehat{v}_\mu(s)$ for all $s \in \mathbb{R}$.*

Proof. See Appendix A. ◻

This immediately gives us the following corollary.

COROLLARY 6.3. *Under the standing assumptions, $\{\mathcal{R}_\mu\}_{\mu>0}$ is an increasing sequence of functionals as $\mu \downarrow 0$, i.e., for every $X \in \mathcal{X}$ we have $\mathcal{R}_\eta(X) \le \mathcal{R}_\mu(X)$ provided $0 < \mu \le \eta$.*

Proof. According to Lemma 6.2.4, for any random variable $X \in \mathcal{X}$, every $t \in \mathbb{R}$, and $\mathbb{P}$-a.a. $\omega \in \Omega$ we have

$$t + \widehat{v}_\eta(X(\omega) - t) \le t + \widehat{v}_\mu(X(\omega) - t)$$

provided $0 < \mu \le \eta$. Consequently, we obtain

$$\mathcal{R}_\eta(X) = \inf_{t\in\mathbb{R}} t + \mathbb{E}[\widehat{v}_\eta(X - t)] \le \inf_{t\in\mathbb{R}} t + \mathbb{E}[\widehat{v}_\mu(X - t)] = \mathcal{R}_\mu(X) \quad (0 < \mu \le \eta). \tag{6.1}$$

Hence, $\{\mathcal{R}_\mu\}$ is an increasing sequence of functionals. ◻

Continuing, for any $X \in \mathcal{X}$ and $\mu > 0$, we define the function $h^X_\mu : \mathbb{R} \to \mathbb{R}$ by

$$h^X_\mu(t) := t + \mathbb{E}[\widehat{v}_\mu(X - t)].$$

LEMMA 6.4. *In addition to the standing assumptions, we consider $\{h^X_\mu\}_{\mu\in(0,C]}$ for some fixed $C > 0$ independent of X. Then $h^X_\mu \xrightarrow{\Gamma} h^X$ given by*

$$h^X(t) := t + \mathbb{E}[v(X - t)].$$

and $\{h^X_\mu\}_{\mu\in(0,C]}$ is equi-coercive, i.e., for all $r \in \mathbb{R}$ there exists a compact subset $K_r \subset \mathbb{R}$ such that $\{t \in \mathbb{R} : h^X_\mu(t) \le r\} \subset K_r$ for all $\mu \in (0, C]$.

REMARK 6.5. *By [15, Proposition 7.7], it suffices to prove the existence of some coercive lower semicontinuous function $\Psi : \mathbb{R} \to \overline{R}$ such that $h^X_\mu \ge \Psi$ for every $\mu \in (0, C]$.*

Proof. As seen in the proof of Corollary 6.3, $\{h^X_\mu\}_{\mu\in(0,C]}$ is an increasing class of functionals as $\mu \downarrow 0$. To see that h^X_μ converges pointwise to h^X, we note that by Lemma 6.2.2 we have $\widehat{v}_\mu(X(\omega)-t) \to v(X(\omega)-t)$ as $\mu \downarrow 0$. Furthermore, Lemma 6.2.3, we have

$$|\widehat{v}_\mu(X(\omega) - t) - \widehat{v}_\mu(0)| = |\widehat{v}_\mu(X(\omega) - t)| \le a_2|X(\omega) - t|$$

Therefore, applying Lebesgue's dominated convergence theorem, we see that h^X_μ converges to h^X pointwise in t. Since v is Lipschitz with constant a_2, we can readily show that h^X is Lipschitz with constant $1 + a_2$ and therefore, lower semicontinuous. Then by Proposition 6.1, $h^X_\mu \xrightarrow{\Gamma} h^X$ as $\mu \downarrow 0$.

Finally, we prove equi-coercivity by demonstrating the existence of a coercive minorant as mentioned in Remark 6.5 above. In the argument below, let $\epsilon \in (0, \min\{a_2 - 1, 1 - a_1\})$. As noted in the proof of Proposition 3.5 and used in the proof of Lemma 6.2, $\widehat{v}_C'$ is strictly monotonically increasing and for $s > 0$, we have $\widehat{v}_C'(s) \in (1, a_2)$. Therefore, by continuity of $\widehat{v}_C'$ and (3.10), there exists some $s_2 > 0$ such that $\widehat{v}_C'(s_2) = a_2 - \epsilon > 1$. Similarly, we can find some $s_1 < 0$, such that $\widehat{v}_C'(s_1) = a_1 + \epsilon < 1$. By convexity, differentiability, and montonicity in μ of $\widehat{v}_\mu$ we have for any $\mu \in (0, C]$:

$$\begin{aligned} \widehat{v}_\mu(s) &\geq \widehat{v}_C(s) \geq \widehat{v}_C(s_1) + (a_1 + \epsilon)(s - s_1) \quad \forall\, s \in \mathbb{R}, \\ \widehat{v}_\mu(s) &\geq \widehat{v}_C(s) \geq \widehat{v}_C(s_2) + (a_2 - \epsilon)(s - s_2) \quad \forall\, s \in \mathbb{R}. \end{aligned}$$

Therefore, it holds that

$$\begin{aligned} &t + \widehat{v}_\mu(X(\omega) - t) \geq \\ &t + \max\{(a_1 + \epsilon)((X(\omega) - t) - s_1) + \widehat{v}_C(s_1), (a_2 - \epsilon)((X(\omega) - t) - s_2) + \widehat{v}_C(s_2)\}, \end{aligned}$$

independently of ω. Consequently, we have

$$\begin{aligned} h_\mu^X(t) \geq \max\{&(1 - (a_1 + \epsilon))t + \widehat{v}_C(s_1) + (a_1 + \epsilon)(\mathbb{E}[X] - s_1), \\ &(1 + \epsilon - a_2)t + \widehat{v}_C(s_2) + (a_2 - \epsilon)(\mathbb{E}[X] - s_2)\}. \end{aligned}$$

Hence, for $|t| \to \infty$ we have $h_\mu^X(t) \to +\infty$. The assertion follows. ∎

Finally, we may combine the results above to prove the main variational convergence result.

THEOREM 6.6. *Under the assumptions of Lemma 6.4, we have* $\mathcal{R}_\mu \xrightarrow{\Gamma} \mathcal{R}$.

Proof. By Corollary 6.3, $\{\mathcal{R}_\mu\}$ is increasing as $\mu \downarrow 0$. Moreover, by [15, Thm. 7.8], the Γ-convergence of h_μ^X to h^X, the equi-coercivity of $\{h_\mu^X\}$, and the definition of the risk measures $\mathcal{R}$, $\mathcal{R}_\mu$ yields the following relation

$$\mathcal{R}(X) = \inf_{t \in \mathbb{R}} t + \mathbb{E}[v(X - t)] = \inf_{t \in \mathbb{R}} h^X(t) = \lim_{\mu \downarrow 0} \inf_{t \in \mathbb{R}} h_\mu^X(t) = \lim_{\mu \downarrow 0} \mathcal{R}_\mu(X).$$

Hence, $\mathcal{R}_\mu \to \mathcal{R}$ pointwise. The assertion then follows from Proposition 6.1. ∎

In light of Theorem 6.6, we can now prove the convergence of approximating minimizers.

THEOREM 6.7. *Let Assumptions 4.1 and 4.2 hold and set $\mathcal{X} = L^1(\Omega, \mathcal{F}, \mathbb{P})$. Furthermore, suppose that $\wp : Z \to \overline{\mathbb{R}}$ is proper, closed, and convex and S is completely continuous. For any sequence $\mu_k \downarrow 0$, suppose that z_k minimizes $f_{\mu_k}(z) := \mathcal{R}_{\mu_k}(\mathcal{J}(S(z))) + \wp(z)$ over $\mathcal{Z}_{\rm ad}$. Then any weak accumlation point of $\{z_k\}$ minimizes $f(z) := \mathcal{R}(\mathcal{J}(S(z))) + \wp(z)$ over $\mathcal{Z}_{\rm ad}$.*

Proof. As argued in the proof of Theorem 6.6, $\mathcal{R}_{\mu_k}$ converges pointwise to $\mathcal{R}$. Moreover, by assumption $F(z) := \mathcal{J}(S(z))$ is completely continuous. Fixing an arbitrary $k \in \mathbb{N}$, we have

$$\mathcal{R}_{\mu_k}(F(z)) + \wp(z) \geq \mathcal{R}_{\mu_k}(F(z_k)) + \wp(z_k)$$

for all $z \in \mathcal{Z}_{\rm ad}$. In light of the complete continuity of F, if $z_{k_j} \rightharpoonup z^\star$ in Z, then $F(z_{k_j}) \to F(z^\star)$ in $\mathcal{X}$. Therefore, it follows from the pointwise and Γ-convergence of

$\mathcal{R}_{\mu_k}$ to $\mathcal{R}$ along with the weak lower-semicontinuity of $\wp$ that:

$$\mathcal{R}(F(z)) + \wp(z) = \lim_{k_j\to\infty} \mathcal{R}_{\mu_{k_j}}(F(z)) + \wp(z) \geq \liminf_{k_j\to\infty} \mathcal{R}_{\mu_{k_j}}[F(z_{k_j})] + \wp(z_{k_j})$$
$$\geq \mathcal{R}[F(z^\star)] + \wp(z^\star)$$

for any $z \in \mathcal{Z}_{\mathrm{ad}}$, as was to be shown. ∎

Remark 6.8. *The complete continuity of S is often guaranteed by the fact that Z is a more regular function space that embeds compactly into the image space of the differential operator. For instance, $Z = L^2(D)$ embeds compactly into $H^{-1}(D)$. Moreover, the existence of weak accumulation points of sequences of solutions can typically be obtained by either the coercivity of $\wp$ or the boundedness of the set $\mathcal{Z}_{\mathrm{ad}}$. Since these are often the situations encountered in PDE-constrained optimization, the additional data assumptions in Theorem 6.7 are arguably mild. In the event that S is not completely continuous, one can still obtain the above result when more structure of $\mathcal{J}$ is available, e.g., when $\mathcal{J}$ is convex with respect to the partial order on $\mathcal{X}$.*

Remark 6.9. *Theorem 6.7 makes no assumptions about the convexity of the optimization problems. However, it is clear that in the non-convex case, the previous results guarantees a certain consistency of the approximation in terms of global solutions only, which may be computationally very difficult to obtain. For convergence of stationary points in the context of a variational smoothing technique for regular measures of risk, we refer the reader to [31].*

7. Implementation and Numerical Results.

We consider the optimal control of an elliptic PDE with uncertain coefficients. For this purpose, let $D \subset \mathbb{R}^n$ be a bounded Lipschitz domain. Let $\kappa \in L^\infty(D \times \Omega)$ be an uncertain coefficient function, which fulfils $\underline{\kappa} \leq \kappa(x,\omega) \leq \overline{\kappa}$ for a.a. $(x,\omega) \in D \times \Omega$ with $0 < \underline{\kappa} \leq \overline{\kappa} < \infty$. We consider the PDE

$$A(\omega)u(\omega) = Bz, \tag{7.1}$$

where $u(\omega) \in H_0^1(D)$ is the state, $z \in L^2(D) = Z$ is the control, and

$$A(\omega) : H_0^1(D) \to H^{-1}(D),\ \langle A(\omega)u, v\rangle_{H^{-1}(D),H_0^1(D)} := \int_D \kappa(x,\omega)\nabla u \cdot \nabla v \,\mathrm{d}x$$
$$B : L^2(D) \to H^{-1}(D),\ \langle Bz, v\rangle_{H^{-1}(D),H_0^1(D)} := \int_D zv \,\mathrm{d}x.$$

Under the assumption on κ, $A(\omega)$ is uniformly elliptic and (7.1) has a unique solution $S(z)(\omega) = A(\omega)^{-1}Bz$ for a.a. $\omega \in \Omega$. In particular, we have $S(z) \in L^\infty(\Omega; H_0^1(D))$. Inserting it into a tracking functional, we have

$$g(z)(\omega) := \tfrac{1}{2}\|\iota S(z)(\omega) - \hat{q}\|^2_{L^2(D)},$$

with the embedding $\iota : H_0^1(D) \hookrightarrow L^2(D)$ and the desired state $\hat{q} \in L^2(D)$. We conclude that $g(z) \in L^\infty(\Omega)$ so that the theory from section 5 is applicable. We let $\gamma > 0$ and set $\wp(z) = \frac{\gamma}{2}\|z\|_Z^2$. Therefore, since g is convex, Lemma 5.2 can be applied, see Remark 5.4.

We compute the derivatives of g by the adjoint approach and discretize the problem by linear finite elements (for D) and Monte Carlo (for Ω). To speed up the evaluation of the samples of g and ∇g, we use a rather exact surrogate model in which the state and adjoint equation are solved by a polynomial chaos discretization in tensor product form with a suitable low-rank tensor solver for the discretized system, see [24]. The objective function and its gradient are computed using efficient low-rank tensor calculus. The required quantities are then sampled from the tensors in parallel. The remaining computations are done with the sampled quantities. We approximate the Hessian by the reference operator: $\nabla^2 g(z)(\omega) \approx \nabla^2 g(z)(\bar{\omega})$, where $\bar{\omega} := \int_\Omega \omega \, \mathrm{d}\mathbb{P}(\omega)$.

We initialize the algorithm with the risk-neutral control z^0, i.e., the solution of (1.1) using $\mathcal{R} \equiv \mathbb{E}$, which is computed by a Newton-CG method using low-rank tensor computations as in [24]. Additonally, we choose $t_0 = \mathbb{E}[g(z^0)]$, $\mu_0 \geq \mu > 0$ ($\mu_0 = 10$ in our tests), and compute W_0, ν_1^0, ν_2^0 from (5.1c), (5.1d), (5.1e). In each iteration the Newton step that solves (5.2) is computed approximately. In our experiments, we found that solving the reduced version (5.4) by CG yields the best computing time and most accurate results. We stopped the CG iteration whenever the relative residual fell below 10^{-2}; using 10^{-8} yielded only slight decreases in the overall iteration counts but actually required more CPU time. The variables z and t are updated using the Newton steps, and the updated auxiliary variables W, ν_1, and ν_2 are computed so that they solve (5.1c), (5.1d), (5.1e), which ensures uniform positivity. Additionally, this procedure is equivalent to applying Newton's method to a reduced problem, namely

$$\min_{z \in Z, t \in \mathbb{R}} t + \mathbb{E}[v_{\mu_k}(g(z) - t)] + \wp(z),$$

see Remark 5.3. We update $\mu_{k+1} = \max\{\mu_{\mathrm{fac}}\, \mu_k, \mu\}$, with $\mu_{\mathrm{fac}} \in (0, 1)$ ($\mu_{\mathrm{fac}} = 0.5$ in our tests). The algorithm is stopped if $\mu_k = \mu$ and the norm of the optimality system residual is below 10^{-4}.

We set $\Omega = (-1, 1)^d$, $d \in \mathbb{N}$, equipped with the uniform distribution, $D = (-1, 1)^2 \subset \mathbb{R}^2$, and $\kappa(x, \omega) = 1 + \sum_{i=1}^d \omega_i \eta_i 1_{D_i}(x)$ with $\eta_i \in (0, 1)$ and the subdomains $D_i \subset D$ covering the domain D. More concretely, the D_i are vertical strips of the same size and $\eta_i = \eta_{\min} + \frac{i-1}{d-1}(\eta_{\max} - \eta_{\min})$, i.e., the deviation in the coefficient increases from left to right. In our tests, we have $d = 4$, $\eta_{\min} = 0.4$, and $\eta_{\max} = 0.7$. Our implementation could be easily adapted for larger d, which would only result in longer runtimes for the tensor computations and sampling. The desired state is $\hat{q}(x) = 1$, and we have 16641 FE nodes and 20000 Monte Carlo samples.

We perform different tests, in which we vary one of the parameters μ (log-barrier parameter), β (quantile parameter), and λ (convex combination paramter). We start with $\mu \in \{0.1, 0.01, 0.001\}$, $\beta = 0.95$, and $\lambda = 1.0$ to investigate the influence of the log-barrier parameter in this setting. Since the difference in the resulting cumulative distribution functions (CDFs) of the random variable objective $g(z^*) + \wp(z^*)$ obtained with $\mu = 0.01$ and $\mu = 0.001$ is hardly recognizable, we proceed with $\mu = 0.001$ in the following tests and do not decrease the log-barrier parameter further.

Figure 1 shows the CDFs of $g(z^\star)(\cdot)$ for different values of β with $\lambda = 1.0$ and $\mu = 0.001$, i.e., we minimize a smoothed version of CVaR_β. In this plot, the expected value is marked by "$\ast$", $\mathrm{CVaR}_{0.5}$ and $\mathrm{CVaR}_{0.9}$ by "+", and $\mathrm{CVaR}_{0.8}$ and $\mathrm{CVaR}_{0.95}$ by "×". As expected, the cheaper deterministic and risk-neutral controls yield better α-quantiles for $\alpha < 0.75$. However, the risk-averse controls clearly dominate for higher values of α relating to the tail. Thus, $g(z^\star)(\cdot)$ in the risk-averse cases is expected to be markedly smaller than the risk-neutral/deterministic for tail events.

Finally, Table 1 shows the number of iterations, computing time, and time spent for solving PDEs with a low-rank tensor method, sampling from tensors in parallel, and solution of the Newton system as well as the required CG iterations. Since we

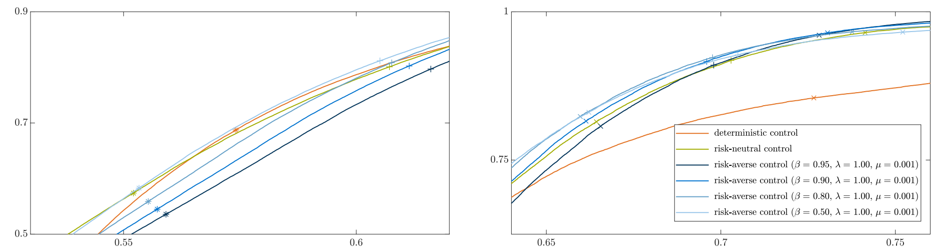

Fig. 1: Cumulative distribution function of the random variable objective function for different optimal controls with $\beta \in \{0.95, 0.9, 0.8, 0.5\}$.

β (CVaR quantile parameter):	0.5	0.8	0.9	0.95
number of iterations (updates of the initial control):	17	18	18	31
computing time (total, in minutes):	8.9	9.4	9.3	16.1
time spent for low-rank tensor computations:	46.5%	45.9%	46.2%	47.2%
time spent for sampling from low-rank tensors:	47.0%	48.6%	47.7%	46.7%
time spent for solution of Newton system:	5.4%	4.4%	5.0%	5.0%
average number of CG iterations (Newton system):	2.9	1.7	1.9	2.0

Table 1: Computing times and statistics for different values of β.

are always solving similar PDEs, similar tensor ranks are sufficient for the desired accuracy and the amount of time spent for the low-rank tensor computations and sampling is rather the same for all tested values of β. Furthermore, the CG method for solving the Newton system performs always comparably well. The total number of iterations is only increased in the case $\beta = 0.95$. Here, the constant approximation of the Hessian $\nabla^2 g(z)$ in $\nabla^2 h(z, \nu_1, \nu 2)$ (see Lemma 5.2) seems to yield worse directions so that reaching the region of fast convergence is harder.

REFERENCES

[1] A. Alexanderian, N. Petra, G. Stadler, and O. Ghattas, *Mean-variance risk-averse optimal control of systems governed by PDEs with random parameter fields using quadratic approximations*, arXiv preprint arXiv:1602.07592, (2016).

[2] J. Appell and P. P. Zabrejko, *Nonlinear Superposition Operators*, Cambridge Tracts in Mathematics, Cambridge University Press, 2008.

[3] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath, *Coherent measures of risk*, Math. Finance, 9 (1999), pp. 203–228.

[4] A. Ben-Tal and M. Teboulle, *Expected utility, penalty functions, and duality in stochastic nonlinear programming*, Management Science, 32 (1986), pp. 1445–1466.

[5] ———, *An old-new concept of convex risk measures: The optimized certainty equivalent*, Mathematical Finance, 17 (2007), pp. 449–476.

[6] P. Benner, S. Dolgov, A. Onwunta, and M. Stoll, *Low-rank solvers for unsteady stokes-brinkman optimal control problem with random data*, Computer Methods in Applied Mechanics and Engineering, 304 (2016), pp. 26 – 54.

[7] J. R. Birge and F. Louveaux, *Introduction to stochastic programming*, Springer-Verlag, New York, 1997.

[8] L. Bonfiglio, P. Perdikaris, S. Brizzolara, and G. Karniadakis, *A multi-fidelity framework for investigating the performance of super-cavitating hydrofoils under uncertain flow*

conditions, in 19th AIAA Non-Deterministic Approaches Conference, American Institute of Aeronautics and Astronautics, 2017, pp. 1–13.

[9] J. F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer Series in Operations Research, Springer-Verlag, New York, 2000.

[10] A. Borzì, V. Schulz, C. Schillings, and G. von Winckel, *On the treatment of distributed uncertainties in PDE constrained optimization*, GAMM Mitteilungen, 33 (2010), pp. 230–246.

[11] A. Borzì and G. von Winckel, *Multigrid methods and sparse-grid collocation techniques for parabolic optimal control problems with random coefficients*, SIAM J. Sci. Comput., 31 (2009), pp. 2172–2192.

[12] ———, *A POD framework to determine robust controls in PDE optimization*, Comput. Vis. Sci., 14 (2011), pp. 91–103.

[13] P. Chen and A. Quarteroni, *Weighted reduced basis method for stochastic optimal control problems with elliptic PDE constraint*, SIAM/ASA Journal on Uncertainty Quantification, 2 (2014), pp. 364–396.

[14] P. Chen, A. Quarteroni, and G. Rozza, *Stochastic optimal robin boundary control problems of advection-dominated elliptic equations*, SIAM Journal on Numerical Analysis, 51 (2013), pp. 2700–2722.

[15] G. Dal Maso, *An Introduction to* Γ*-Convergence*, vol. 8 of Progress in Nonlinear Differential Equations and their Applications, Birkhäuser Boston, Inc., Boston, MA, 1993.

[16] D. Dentcheva and A. Ruszczyński, *Optimization with stochastic dominance constraints*, SIAM J. Optim., 14 (2003), pp. 548–566.

[17] J. Dupačová, *The minimax approach to stochastic programming and an illustrative application*, Stochastics, 20 (1987), pp. 73–88.

[18] L. J. Durlofsky and Y. Chen, *Uncertainty Quantification for Subsurface Flow Problems Using Coarse-Scale Models*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, ch. 6, pp. 163–202.

[19] I. Ekeland and R. Témam, *Convex Analysis and Variational Problems*, vol. 28 of Classics in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, english ed., 1999. Translated from the French.

[20] M. S. Eldred, *Design under uncertainty employing stochastic expansion methods*, Int. J. Uncertain. Quantif., 1 (2011), pp. 119–146.

[21] M. S. Eldred, H. Agarwal, V. M. Perez, S. F. Wojtkiewicz, Jr., and J. E. Renaud, *Investigation of reliability method formulations in DAKOTA/UQ*, Struct. Infrastruct. Eng.: Maint., Man., Life-Cycle Des. Perform., 3 (2007), pp. 199–213.

[22] M. S. Eldred and B. J. Bichon, *Second-order reliability formulations in DAKOTA/UQ*, Proceedings of 47th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference, Newport, RI, May 1–4, (2006). Paper No. AIAA-2006-1828.

[23] S. Garreis, *Optimal Control under Uncertainty: Theory and Numerical Solution with Low-Rank Tensors*, PhD thesis, Technical University of Munich, 2019.

[24] S. Garreis and M. Ulbrich, *Constrained optimization with low-rank tensors and applications to parametric problems with PDEs*, SIAM Journal on Scientific Computing, 39 (2017), pp. A25–A54.

[25] E. Hille and R. S. Phillips, *Functional analysis and semi-groups*, American Mathematical Society Colloquium Publications, vol. 31, American Mathematical Society, Providence, R. I., 1957. rev. ed.

[26] M. Hintermüller, K. Ito, and K. Kunisch, *The primal-dual active set strategy as a semismooth Newton method*, SIAM Journal on Optimization, 13 (2002), pp. 865–888.

[27] P. Kolvenbach, O. Lass, and S. Ulbrich, *An approach for robust PDE-constrained optimization with application to shape optimization of electrical engines and of dynamic elastic structures under uncertainty*, Optim. Eng., 19 (2018), pp. 697–731.

[28] D. P. Kouri, M. Heinkenschloss, D. Ridzal, and B. G. van Bloemen Waanders, *A trust-region algorithm with adaptive stochastic collocation for PDE optimization under uncertainty*, SIAM Journal on Scientific Computing, 35 (2013), pp. A1847–A1879.

[29] ———, *Inexact objective function evaluations in a trust-region algorithm for PDE-constrained optimization under uncertainty*, SIAM Journal on Scientific Computing, 36 (2014), pp. A3011–A3029.

[30] D. P. Kouri and T. M. Surowiec, *Risk-averse PDE-constrained optimization using the conditional value-at-risk*, SIAM Journal on Optimization, 26 (2016), pp. 365–396.

[31] ———, *Epi-regularization of risk measures for PDE-constrained optimization*, (2017). submitted to Mathematics of Operations Research.

[32] ———, *Existence and optimality conditions for risk-averse PDE-constrained optimization*,

SIAM/ASA Journal on Uncertainty Quantification, 6 (2018), pp. 787–815.
[33] ———, *Risk-averse optimal control of semilinear elliptic PDEs*, (2018). submitted to Control, Optimisation and Calculus of Variations.
[34] O. Lass and S. Ulbrich, *Model order reduction techniques with a posteriori error control for nonlinear robust optimization governed by partial differential equations*, SIAM J. Sci. Comput., 39 (2017), pp. S112–S139.
[35] N. Petra, H. Zhu, G. Stadler, T. J. R. Hughes, and O. Ghattas, *An inexact Gauss-Newton method for inversion of basal sliding and rheology parameters in a nonlinear Stokes ice sheet model*, Journal of Glaciology, 58 (2012).
[36] A. Prékopa, *Stochastic Programming*, vol. 324 of Mathematics and its Applications, Kluwer Academic Publishers Group, Dordrecht, 1995.
[37] ———, *Probabilistic programming*, in Stochastic Programming, vol. 10 of Handbooks Oper. Res. Management Sci., Elsevier Sci. B. V., Amsterdam, 2003, pp. 267–351.
[38] R. T. Rockafellar and S. Uryasev, *Optimization of conditional value-at-risk*, Journal of Risk, 2 (2000), pp. 21–41.
[39] R. T. Rockafellar and S. Uryasev, *The fundamental risk quadrangle in risk management, optimization and statistical estimation*, Surveys in Operations Research and Management Science, 18 (2013), pp. 33–53.
[40] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Springer Verlag, Berlin, Heidelberg, New York, 1998.
[41] E. Rosseel and G. N. Wells, *Optimal control with stochastic PDE constraints and uncertain controls*, Computer Methods in Applied Mechanics and Engineering, 213?Ai216 (2012), pp. 152–167.
[42] A. Ruszczyński and A. Shapiro, *Optimization of convex risk functions*, Math. Oper. Res., 31 (2006), pp. 433–452. see also [43].
[43] ———, *Corrigendum to: "Optimization of convex risk functions" [Math. Oper. Res.* **31** *(2006), no. 3, 433–452]*, Math. Oper. Res., 32 (2007), p. 496.
[44] A. Schiela, *Barrier methods for optimal control problems with state constraints*, SIAM Journal on Optimization, 20 (2009), pp. 1002–1031.
[45] A. Shapiro and S. Ahmed, *On a class of minimax stochastic programs*, SIAM Journal on Optimization, 14 (2004), pp. 1237–1249.
[46] A. Shapiro, D. Dentcheva, and A. Ruszczynski, *Lectures on Stochastic Programming: Modeling and Theory, Second Edition*, MOS-SIAM Series on Optimization, Society for Industrial and Applied Mathematics, Philadelphia, 2014.
[47] A. Shapiro and A. Kleywegt, *Minimax analysis of stochastic problems*, Optimization Methods and Software, 17 (2002), pp. 523–542.
[48] G. Stadler, M. Gurnis, C. Burstedde, L. C. Wilcox, L. Alisic, and O. Ghattas, *The dynamics of plate tectonics and mantle flow: From local to global scales*, Science, 329 (2010), pp. 1033–1038.
[49] D. M. Tartakovsky, A. Guadagnini, and M. Riva, *Stochastic averaging of nonlinear flows in heterogeneous porous media*, Journal of Fluid Mechanics, 492 (2003), pp. 47–62.
[50] H. Tiesler, R. M. Kirby, D. Xiu, and T. Preusser, *Stochastic collocation for optimal control problems with stochastic PDE constraints*, SIAM Journal on Control and Optimization, 50 (2012), pp. 2659–2682.
[51] M. Ulbrich, *Semismooth Newton methods for operator equations in function spaces*, SIAM Journal on Optimization, 13 (2002), pp. 805–841.
[52] ———, *Semismooth Newton Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces*, vol. 11 of MOS-SIAM Series on Optimization, SIAM; MOS, Philadelphia, PA, 2011.
[53] M. Ulbrich and S. Ulbrich, *Primal-dual interior-point methods for PDE-constrained optimization*, Mathematical Programming, 117 (2009), pp. 435–485.
[54] M. Weiser and A. Schiela, *Function space interior point methods for PDE constrained optimization*, PAMM, 4 (2004), pp. 43–46.

Appendix A. Proof of Lemma 6.2.

Proof. Statement 3. follows immediately from (3.7).

For 1., we start by noting that $v(0) = 0 = \widehat{v}_\mu(0)$ and $\widehat{v}'_\mu(0) = 1$, see (3.13). Moreover, for $s > 0$, we have $\widehat{v}'_\mu(s) \in (1, a_2)$, see the derivation of (3.11). Therefore,

$$\widehat{v}_\mu(s) = \int_0^s \widehat{v}'_\mu(\tau) d\tau < a_2 s = v(s).$$

This follows analogously for the case when $s < 0$, using in part the fact that $\widehat{v}'_\mu(s) \in (a_1, 1)$.

In order to prove 2., we need several arguments. We recall that

$$w_\mu(s) = \mu + \tfrac{1}{2}(a_1 + a_2)s + \tfrac{1}{2}\sqrt{(a_2 - a_1)^2 s^2 + 4\mu^2}$$

and observe that

$$\lim_{\mu\to 0^+} w_\mu(s) = \tfrac{1}{2}(a_1 + a_2)s + \tfrac{1}{2}|a_2 - a_1|\,|s| = \max\{a_1 s, a_2 s\}.$$

follows from $a_1 < a_2$ and considering $s \le 0$ and $s \ge 0$ separately. Furthermore, we consider the limit $\lim_{\mu\to 0^+} \mu \cdot \ln(w_\mu(s) - a_1 s)$. We use

$$\lim_{\mu\to 0^+} w_\mu(s) - a_1 s = \tfrac{1}{2}(a_2 - a_1)s + \tfrac{1}{2}|a_2 - a_1|\,|s| = (a_2 - a_1)\max\{0, s\}.$$

For $s > 0$ it follows that $\lim_{\mu\to 0^+} \mu \cdot \ln(w_\mu(s) - a_1 s) = 0$. In the case $s = 0$, we have $\lim_{\mu\to 0^+} \mu \cdot \ln(w_\mu(s) - a_1 s) = \lim_{\mu\to 0^+} \mu \cdot \ln(2\mu) = 0$. For $s < 0$ we get

$$\begin{aligned}
&\lim_{\mu\to 0^+} \mu \ln(w_\mu(s) - a_1 s)\\
&= \lim_{\mu\to 0^+} \mu \ln\big(\mu + \tfrac{a_2-a_1}{2}s + \tfrac{1}{2}|a_2 - a_1||s| + \tfrac{\mu^2}{|a_2-a_1||s|} + o(\mu^2)\big) =\\
&= \lim_{\mu\to 0^+} \mu \ln\big(\mu + \tfrac{\mu^2}{(a_2-a_1)|s|} + o(\mu^2)\big) = 0
\end{aligned}$$

Summarizing, we have $\lim_{\mu\to 0^+} \mu \cdot \ln(w_\mu(s) - a_1 s) = 0$ for all s. Analogously, it follows that $\lim_{\mu\to 0^+} \mu \cdot \ln(w_\mu(s) - a_2 s) = 0$.

Next, we see that

$$\lim_{\mu\to 0^+} \zeta(\mu) = \lim_{\mu\to 0^+} \mu\big(\ln\big(\tfrac{a_2-a_1}{a_2-1}\mu\big) + \ln\big(\tfrac{a_2-a_1}{1-a_1}\mu\big) - 2\big) = 0$$

holds. Finally, we have $\lim_{\mu\to 0^+} v_\mu(s) = v(s)$ for all $s \in \mathbb{R}$ and hence $\lim_{\mu\to 0^+} \widehat{v}_\mu(s) = v(s)$ as well due to $d(\mu) \to 0$ as $\mu \to 0$.

In order to prove 4., we investigate the sign properties of the derivatives of v_μ as a function of $\mu > 0$ for fixed $s \in \mathbb{R}$. We start by observing that

$$\partial_\mu v_\mu(s) = \partial_\mu w_\mu(s) - \ln(w_\mu(s) - a_1 s) - \mu\,\tfrac{\partial_\mu w_\mu(s)}{w_\mu(s) - a_1 s} - \ln(w_\mu(s) - a_2 s) - \mu\,\tfrac{\partial_\mu w_\mu(s)}{w_\mu(s) - a_2 s} + \zeta'(\mu),$$

where $\partial_\mu w_\mu(s) = 1 + \frac{2\mu}{\sqrt{(a_2-a_1)^2 s^2 + 4\mu^2}}$ and $\zeta'(\mu) = \ln(\frac{a_2-a_1}{a_2-1}\mu) + \ln(\frac{a_2-a_1}{1-a_1}\mu)$. Next, writing $\beta = a_2 - a_1 > 0$, $\gamma = \sqrt{\beta^2 s^2 + 4\mu^2} > 0$ and $\alpha_+ = \frac{\beta s + \gamma}{2}$, $\alpha_- = \frac{-\beta s + \gamma}{2}$, we have

$$\begin{aligned}
w_\mu(s) - a_1 s &= \mu + \alpha_+,\\
w_\mu(s) - a_2 s &= \mu + \alpha_-,\\
(w_\mu(s) - a_1 s)(w_\mu(s) - a_2 s) &= (\mu + \alpha_+)(\mu + \alpha_-) = \mu(2\mu + \gamma),\\
\partial_\mu w_\mu(s) &= \tfrac{2\mu+\gamma}{\gamma}.
\end{aligned}$$

By substitution, the derivative of v_μ becomes

$$\begin{aligned}
\partial_\mu v_\mu(s) &= \tfrac{2\mu+\gamma}{\gamma} - \mu\tfrac{\gamma+2\mu}{\gamma(\mu+\alpha_+)} - \mu\tfrac{\gamma+2\mu}{\gamma(\mu+\alpha_-)} - \ln\big(\mu(2\mu+\gamma)\big) + \ln\big(\tfrac{\beta^2}{(1-a_1)(a_2-1)}\mu^2\big)\\
&= \tfrac{\mu(2\mu+\gamma)^2 - \mu(\gamma+2\mu)(2\mu+\alpha_- + \alpha_+)}{\gamma(\mu+\alpha_+)(\mu+\alpha_-)} + \ln\big(\tfrac{\beta^2\mu}{(1-a_1)(a_2-1)(2\mu+\gamma)}\big)\\
&= \ln\Big(\tfrac{(a_2-a_1)^2\mu}{(1-a_1)(a_2-1)\big(2\mu + \sqrt{(a_2-a_1)^2 s^2 + 4\mu^2}\big)}\Big).
\end{aligned}$$

Now consider $\widehat{v}_\mu(s) = v_\mu(s + d(\mu)) - d(\mu)$. Then,

$$\partial_\mu \widehat{v}_\mu(s) = v'_\mu(s + d(\mu))d'(\mu) + \partial_\mu v_\mu(s + d(\mu)) - d'(\mu),$$

where $v'_\mu(s) = w'_\mu(s) - \mu \frac{w'_\mu(s) - a_1}{w_\mu(s) - a_1 s} - \mu \frac{w'_\mu(s) - a_2}{w_\mu(s) - a_2 s}$ with $w'_\mu(s) = \frac{a_1 + a_2}{2} + \frac{\beta^2 s}{2\gamma}$. We simplify

$$\begin{aligned}
v'_\mu(s) &= \frac{a_1+a_2}{2} + \frac{\beta^2 s}{2\gamma} - \mu \frac{\frac{\beta}{2} + \frac{\beta^2 s}{2\gamma}}{\mu + \alpha_+} - \mu \frac{-\frac{\beta}{2} + \frac{\beta^2 s}{2\gamma}}{\mu + \alpha_-} \\
&= \frac{a_1+a_2}{2} + \frac{\beta^2 s}{2\gamma} - \mu \left(\frac{\beta}{2} \left(\frac{1}{\mu+\alpha_+} - \frac{1}{\mu+\alpha_-} \right) + \frac{\beta^2 s}{2\gamma} \left(\frac{1}{\mu+\alpha_+} + \frac{1}{\mu+\alpha_-} \right) \right) \\
&= \frac{a_1+a_2}{2} + \frac{\beta^2 s}{2\gamma} - \mu \left(\frac{\beta}{2} \frac{\alpha_- - \alpha_+}{\mu(2\mu+\gamma)} + \frac{\beta^2 s}{2\gamma} \frac{2\mu + \alpha_- + \alpha_+}{\mu(2\mu+\gamma)} \right) = \frac{a_1+a_2}{2} + \frac{\beta^2 s}{4\mu + 2\gamma}
\end{aligned}$$

Now, writing $\tilde{s} = s + d(\mu)$, $\tilde{\gamma} = \sqrt{\beta^2 \tilde{s}^2 + 4\mu^2}$, $\rho_1 = 1 - a_1 > 0$, and $\rho_2 = a_2 - 1 > 0$ so that $d(\mu) = \frac{\rho_1 - \rho_2}{\rho_1 \rho_2} \mu =: \kappa\mu$, we get

$$\begin{aligned}
\partial_\mu \widehat{v}_\mu(s) &= \left(\frac{a_1+a_2}{2} + \frac{\beta^2 \tilde{s}}{4\mu + 2\tilde{\gamma}} - 1 \right) \kappa + \ln\left(\frac{\beta^2 \mu}{\rho_1 \rho_2 \left(2\mu + \tilde{\gamma}\right)} \right) \\
&= \frac{\kappa}{2} \left(\rho_2 - \rho_1 + \frac{\beta^2 \tilde{s}}{2\mu + \tilde{\gamma}} \right) + \ln\left(\frac{\beta^2 \mu}{\rho_1 \rho_2 \left(2\mu + \tilde{\gamma}\right)} \right).
\end{aligned}$$

We compute

$$\begin{aligned}
\lim_{\mu \to +\infty} \partial_\mu \widehat{v}_\mu(s) &= \frac{\kappa}{2} \left(\rho_2 - \rho_1 + \frac{\beta^2 \kappa}{2 + \sqrt{\beta^2 \kappa^2 + 4}} \right) + \ln\left(\frac{\beta^2}{\rho_1 \rho_2 (2 + \sqrt{\beta^2 \kappa^2 + 4})} \right) \\
&= \frac{\kappa}{2} \left(\rho_2 - \rho_1 + \frac{\beta^2 \kappa \rho_1 \rho_2}{\beta^2} \right) + \ln(1) = 0.
\end{aligned} \tag{A.1}$$

We have used that $\beta = \rho_1 + \rho_2$ and thus

$$2 + \sqrt{\beta^2 \kappa^2 + 4} = 2 + \sqrt{\frac{(\rho_1+\rho_2)^2 (\rho_1 - \rho_2)^2 + 4\rho_1^2 \rho_2^2}{\rho_1^2 \rho_2^2}} = 2 + \frac{\rho_1^2 + \rho_2^2}{\rho_1 \rho_2} = \frac{\beta^2}{\rho_1 \rho_2}.$$

The second derivative is

$$\begin{aligned}
\partial^2_{\mu\mu} \widehat{v}_\mu(s) &= \frac{\kappa}{2} \frac{(2\mu+\tilde{\gamma})\beta^2 \kappa - \beta^2 \tilde{s}(2 + \tilde{\gamma}')}{(2\mu+\tilde{\gamma})^2} + \frac{\rho_1 \rho_2 (2\mu + \tilde{\gamma})}{\beta^2 \mu} \frac{\rho_1 \rho_2 (2\mu+\tilde{\gamma})\beta^2 - \beta^2 \mu \rho_1 \rho_2 (2 + \tilde{\gamma}')}{\rho_1^2 \rho_2^2 (2\mu+\tilde{\gamma})^2} \\
&= \frac{(2\mu+\tilde{\gamma})\beta^2 \kappa^2 - \beta^2 \tilde{s}(2+\tilde{\gamma}')\kappa}{2(2\mu+\tilde{\gamma})^2} + \frac{(2\mu+\tilde{\gamma}) - \mu(2+\tilde{\gamma}')}{\mu(2\mu+\tilde{\gamma})} \\
&= \frac{\mu(2\mu+\tilde{\gamma})\beta^2\kappa^2 - \mu\beta^2 \tilde{s}(2+\tilde{\gamma}')\kappa + 2(2\mu+\tilde{\gamma})^2 - 2\mu(2\mu+\tilde{\gamma})(2+\tilde{\gamma}')}{2\mu(2\mu+\tilde{\gamma})^2}
\end{aligned}$$

with $\tilde{\gamma}' = \frac{\beta^2 \tilde{s}\kappa + 4\mu}{\sqrt{\beta^2 \tilde{s}^2 + 4\mu^2}} = \frac{\beta^2 \kappa \tilde{s} + 4\mu}{\tilde{\gamma}}$. The numerator is

$$\begin{aligned}
&2\mu^2 \beta^2 \kappa^2 + \mu \beta^2 \kappa^2 \tilde{\gamma} - 2\mu\beta^2 \kappa \tilde{s} - \mu \beta^2 \kappa \tilde{s} \tilde{\gamma}' \\
&\quad + 8\mu^2 + 8\mu\tilde{\gamma} + 2\tilde{\gamma}^2 - 8\mu^2 - 4\mu^2 \tilde{\gamma}' - 4\mu\tilde{\gamma} - 2\mu\tilde{\gamma}\tilde{\gamma}' \\
&= 2\mu^2 \beta^2 \kappa^2 + \mu(\beta^2 \kappa^2 + 4)\tilde{\gamma} - 2\mu \beta^2 \kappa \tilde{s} - \mu \frac{(\beta^2 \kappa \tilde{s} + 4\mu)^2}{\tilde{\gamma}} + 2(\beta^2 \tilde{s}^2 + 4\mu^2) - 2\mu(\beta^2 \kappa \tilde{s} + 4\mu) \\
&= 2\mu^2 \beta^2 \kappa^2 + \mu(\beta^2 \kappa^2 + 4)\tilde{\gamma} - 4\mu \beta^2 \kappa (s + \kappa\mu) - \mu \frac{(\beta^2 \kappa \tilde{s} + 4\mu)^2}{\tilde{\gamma}} + 2\beta^2 (s + \kappa\mu)^2 \\
&= 2\beta^2 s^2 + \frac{1}{\tilde{\gamma}} \left(\mu(\beta^2 \kappa^2 + 4)(\beta^2 \tilde{s}^2 + 4\mu^2) - \mu(\beta^2 \kappa \tilde{s} + 4\mu)^2 \right) \\
&= 2\beta^2 s^2 + \frac{\mu}{\tilde{\gamma}} \left(\beta^4 \kappa^2 \tilde{s}^2 + 4\mu^2 \beta^2 \kappa^2 + 4\beta^2 \tilde{s}^2 + 16\mu^2 - \beta^4 \kappa^2 \tilde{s}^2 - 8\mu\beta^2 \kappa \tilde{s} - 16\mu^2 \right) \\
&= 2\beta^2 s^2 + \frac{\mu}{\tilde{\gamma}} \left(4\mu^2 \beta^2 \kappa^2 + 4\beta^2 \tilde{s}^2 - 8\mu\beta^2 \kappa \tilde{s} \right) = 2\beta^2 s^2 + \frac{4\mu}{\tilde{\gamma}} \left(\mu\beta\kappa - \beta\tilde{s} \right)^2 \geq 0.
\end{aligned}$$

Therefore, having $2\mu(2\mu + \tilde{\gamma})^2 > 0$, $\partial^2_{\mu\mu} \widehat{v}_\mu(s) \geq 0$ holds for all $s \in \mathbb{R}$ and $\mu > 0$ so that $\partial_\mu \widehat{v}_\mu(s)$ is increasing w.r.t. μ. Hence, together with (A.1), $\partial_\mu \widehat{v}_\mu(s) \leq 0$ follows for all $s \in \mathbb{R}$, $\mu > 0$. This completes the proof. ∎