AN APPROXIMATION SCHEME FOR DISTRIBUTIONALLY ROBUST PDE-CONSTRAINED OPTIMIZATION

JOHANNES MILZ* AND MICHAEL ULBRICH*

Abstract. We develop a sampling-free approximation scheme for distributionally robust PDE-constrained optimization problems, which are min-max control problems. We define the ambiguity set through moment and entropic constraints. We use second-order Taylor's expansions of the reduced objective function w.r.t. uncertain parameters, allowing us to compute the expected value of the quadratic function explicitly. The objective function of the approximated min-max problem separates into a trust-region problem and a semidefinite program. We construct smoothing functions for the optimal value functions defined by these problems. We prove the existence of optimal solutions for the distributionally robust control problem, and the approximated and smoothed problems, and show that a worst-case distribution exists. For the numerical solution of the approximated problem, we develop a homotopy method that computes a sequence of stationary points of smoothed problems while decreasing smoothing parameters to zero. The adjoint approach is used to compute derivatives of the smoothing functions. Numerical results for two nonlinear optimization problems are presented.

Key words. PDE-constrained optimization under uncertainty, distributionally robust optimization, trust-region problem, smoothing functions, smoothing methods

AMS subject classifications. 49J20, 49J35, 49J55, 65K05, 90C15, 90C26, 90C59

1. Introduction. Stochastic programming offers an approach to optimization under uncertainty. In order to apply this approach, uncertain parameters are modeled as a random vector distributed according to a probability distribution, which is often unknown and approximated by a known nominal probability measure. Distributionally robust optimization (DRO) allows the decision maker to take into account the "uncertainty" of that nominal probability distribution; the goal is to find an optimal solution that is resilient to distributional uncertainty. The task is formulated as the minimization of the worst-case expected value of a parameterized objective function. The worst-case is computed w.r.t. probability measures contained in an ambiguity set, which models the distributional uncertainty. These sets can be defined through moment constraints [18, 52] and/or various probability distances [19, 41, 44]. Certain classes of DRO problems (DROPs) can be solved efficiently [18, 19, 52]. However, when the parametrized objective function is nonconcave as a function of the parameters or implicitly defined, DROPs are generally intractable [18, 52]. We develop a computationally tractable sampling-free approximation scheme, based on [38], for the distributionally robust nonlinear, partial differential equation (PDE)-constrained optimal control problem

$$\min_{u\in U_{\mathrm{ad}}} \Big\{ \sup_{P\in\mathcal{P}} \mathbb{E}_P[J(S(u,\xi),u,\xi)] \Big\}, \tag{1.1}$$

where $\mathbb{E}_P[J(S(u,\xi),u,\xi)] = \int_{\mathbb{R}^p} J(S(u,\xi),u,\xi)dP(\xi)$, $U_{\mathrm{ad}} \subset U$ is the set of admissible controls, and U is a Hilbert space. Moreover, $J : Y \times U \times \mathbb{R}^p \to \mathbb{R}$ is the parametrized cost function, and $S : U \times \mathbb{R}^p \to Y$ is the solution operator of the parameterized PDE $e(S(u,\xi),u,\xi) = 0$ for $\xi \in \mathbb{R}^p$, where $e : Y \times U \times \mathbb{R}^p \to Z$, and Y, Z are Banach spaces. Motivated by [14, 18, 38, 47], we use moment and entropic constraints to

*Technical University of Munich, Chair of Mathematical Optimization, Department of Mathematics, Boltzmannstr. 3, 85748 Garching, Germany (milz@ma.tum.de, mulbrich@ma.tum.de). The project was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)–project number 188264188/GRK1754.

define the ambiguity set $\mathcal{P}$:

$$\begin{aligned}\mathcal{P} = \{\, P \in \mathcal{M} : \|\bar{\Sigma}^{-1/2}(\mathbb{E}_P[\xi] - \bar{\mu})\|_2 \le \Delta, \quad \sigma_0\bar{\Sigma} \preccurlyeq \mathrm{Cov}_P[\xi] \preccurlyeq \sigma_1\bar{\Sigma},\\ \ln \mathbb{E}_P[\exp(d^T(\xi - \mathbb{E}_P[\xi]))] \le (\sigma_1/2)d^T\bar{\Sigma}d \quad \text{for all} \quad d \in \mathbb{R}^p \,\},\end{aligned} \tag{1.2}$$

where $\Delta > 0$, $\bar{\mu} \in \mathbb{R}^p$, $\bar{\Sigma} \in \mathbb{R}^{p\times p}$ is a symmetric, positive definite matrix, σ_0, $\sigma_1 \in \mathbb{R}_+$ fulfill $\sigma_0 < \sigma_1$, and $\preccurlyeq$ is the Löwner partial order. Here, $\mathbb{E}_P[\xi]$ denotes the expectation and $\mathrm{Cov}_P[\xi]$ the covariance matrix both w.r.t $P \in \mathcal{M}$, and $\mathcal{M}$ is the set of probability measures on $\mathbb{R}^p$ equipped with its Borel-σ-field. The first two conditions in (1.2) model confidence regions for the mean and the covariance of the random vector ξ [38, sect. 2]. The ambiguity set $\mathcal{P}$ contains all distributions of strictly sub-Gaussian random vectors, in particular all normal distributions, with mean μ satisfying $\|\bar{\Sigma}^{-1/2}(\mu - \bar{\mu})\|_2 \le \Delta$ and covariance matrix Σ fulfilling $\sigma_0\bar{\Sigma} \preccurlyeq \Sigma \preccurlyeq \sigma_1\bar{\Sigma}$; cf. [10, pp. 185–186]. The data, such as $\bar{\mu}$ and $\bar{\Sigma}$, used in (1.2) may be defined using empirical estimates, similar to the choices made in [18, sect. 3.4] and [47, sect. 3.3]. We refer to [14, 18, 47, 52] for further discussions of moment-based ambiguity sets, and to [19, 44] for discussions of potential shortcomings of these sets.

Our approximation scheme and algorithmic approach builds on the one developed in [38] for finite dimensional optimization problems, which we extend to control problems with PDEs. We define the *reduced parametrized objective function* $\widehat{J} : U \times \mathbb{R}^p \to \mathbb{R}$ associated to (1.1) by

$$\widehat{J}(u, \xi) = J(S(u, \xi), u, \xi). \tag{1.3}$$

For each $u \in U$, we require that $\widehat{J}(u, \cdot)$ is twice continuously differentiable in a neighborhood of $\bar{\mu}$, and we approximate $\widehat{J}(u, \cdot)$ via the second-order Taylor's expansion

$$Q(u, \xi; \bar{\mu}) = \widehat{J}(u, \bar{\mu}) + \nabla_\xi \widehat{J}(u, \bar{\mu})^T(\xi - \bar{\mu}) + (1/2)(\xi - \bar{\mu})^T\nabla_{\xi\xi}\widehat{J}(u, \bar{\mu})(\xi - \bar{\mu}), \tag{1.4}$$

where $Q(\cdot, \cdot; \bar{\mu}) : U \times \mathbb{R}^p \to \mathbb{R}$. Here, $\nabla_\xi \widehat{J}(u; \cdot)$ and $\nabla_{\xi\xi}\widehat{J}(u; \cdot)$ denotes the gradient and the Hessian of $\widehat{J}(u; \cdot)$, respectively. The structure of the function Q allows us to compute $\mathbb{E}_P[Q(u, \xi; \bar{\mu})]$ for every $u \in U$ explicitly [38, sect. 1].

As in [38, sect. 1], we formulate the computationally tractable approximation of (1.1) via the distributionally robust optimization problem (DROP)

$$\min_{u \in U_{\mathrm{ad}}} \Big\{ \sup_{P \in \mathcal{P}} \mathbb{E}_P[Q(u, \xi; \bar{\mu})] \Big\}. \tag{1.5}$$

From [38, sect. 1], we obtain that the objective function $\Psi : U \to \mathbb{R}$ of (1.5) can be expressed as

$$\Psi(u) = \widehat{J}(u, \bar{\mu}) + \varphi(u) + \psi(u), \tag{1.6}$$

where $\varphi : U \to \mathbb{R}$ is the optimal value function of the semidefinite program (SDP)

$$\varphi(u) = \max_{\Sigma \in \mathbb{S}^p_+} \{\, (1/2)\nabla_{\xi\xi}\widehat{J}(u; \bar{\mu}) \bullet \Sigma : \sigma_0\bar{\Sigma} \preccurlyeq \Sigma \preccurlyeq \sigma_1\bar{\Sigma} \,\}, \tag{1.7}$$

Here, $\bullet$ is the Frobenius inner product. Moreover, $\psi : U \to \mathbb{R}$ is the optimal value function of the trust-region problem (TRP)

$$\psi(u) = \max_{d \in \mathbb{R}^p} \{\, \nabla_\xi \widehat{J}(u; \bar{\mu})^T d + (1/2)d^T\nabla_{\xi\xi}\widehat{J}(u; \bar{\mu})d : \|\bar{\Sigma}^{-1/2}d\|_2 \le \Delta \,\}. \tag{1.8}$$

We show, under suitable assumptions, that the approximation error of the cost function of (1.1) and (1.5) is small. The cost function Ψ of (1.5) can efficiently be evaluated (being the sum of a TRP and an SDP) without further approximations, such as sampling; however it is nonsmooth [38]. For the numerical solution of (1.5), we use the smoothing function of Ψ, that is, a smooth approximation of Ψ with explicit bounds on the smoothing error, which was constructed in [38, sect. 6]. This function defines smoothed DROPs parameterized by smoothing parameters. Deviating from [38], we use a slightly different definition of a smoothing function that is independent of convergence notions. We extend the homotopy method developed in [38, Alg. 3.1] to allow the numerical treatment of infinite dimensional problems. The algorithm applies optimization solvers to smoothed DROPs while decreasing smoothing parameters to zero. All required derivatives of the smoothing function of Ψ are computed using `UFL` [4] and `FEniCs` [3]. Furthermore, we prove the existence of optimal solutions for the DROP (1.1), of the approximated DROP (1.5) and smoothed ones, and the existence of a worst-case distribution in (1.1). Moreover, we present a convergence result of the homotopy method in Hilbert spaces; global convergence for a finite dimensional setting has been established in [38, Thm. 6.2].

A popular solution approach (see, e.g., [18, 52, 14]) for moment-based DROPs, that is, when the ambiguity set $\mathcal{P}$ is defined by moment constraints, exploits the fact that a Lagrangian dual of the maximization problem in (1.1) is a robust optimization (RO) problem [18, sect. 2.1]. Under suitable assumptions, strong duality holds and the dual can be concatenated with the upper-level problem to obtain a single-level problem [18, 52]. The tractability of the dual depends on the structure of $\widehat{J}(u,\cdot)$ [18, 52, 6]. For example, if $\widehat{J}(u,\cdot)$ is the pointwise maximum of affine functions for all $u \in U$, the dual is tractable [18, sect. 4.1]. However, this approach does not result in explicit single-level programs when $\widehat{J}(u,\cdot)$ is implicitly defined or non-quadratic. Moreover, a "robustified" constraint appears in the single-level problem [18, p. 597] and, hence, available solvers for PDE-constrained problems cannot be applied to it.

The DROP (1.1) becomes a risk-neutral problem when the ambiguity set $\mathcal{P}$ is a singleton. We refer to [45] for an overview of stochastic programming. Risk-neutral optimization with PDEs has been considered in, e.g., [20, 31].

The objective function $f : U \to \mathbb{R} \cup \{\infty\}$ of (1.1),

$$f(u) = \sup_{P\in\mathcal{P}} \mathbb{E}_P[\widehat{J}(u,\xi)], \tag{1.9}$$

is convex if $\widehat{J}(\cdot,\xi)$ is convex for all $\xi \in \mathbb{R}^p$ and $\mathbb{E}_P[\widehat{J}(u,\xi)]$ is well-defined for all $P \in \mathcal{P}$. Similarly, the composition of a convex, monotonic risk measure with $\widehat{J}$ is convex if $\widehat{J}(\cdot,\xi)$ is convex for all $\xi \in \mathbb{R}^p$ [45, Prop. 6.8]. Risk-averse optimization with coherent risk measures can equivalently be reformulated as min-max problems similar to that in (1.1) where the ambiguity set is the domain of the convex conjugate of the risk measure [42, 45, 34]. In the literature on optimization with PDEs, popular risk measures are: the conditional value-at-risk [32, 33], the entropic risk measure [34] and the mean-plus-variance risk measure [1, 12, 48]. Risk-neutral and risk-averse shape optimization are investigated in [15, 16]. Existence results for solutions and optimality conditions of risk-averse control problems with PDEs are provided in [32, 33].

DRO with PDEs has been considered in [29] with ambiguity sets including the support constraint $P(\xi \in \Xi) = 1$, where $\Xi \subset \mathbb{R}^p$ is a convex, compact Lipschitz domain. An inner approximation of the ambiguity set is constructed via a measure discretization, error bounds are derived and the author has shown that the objective function of the min-max problem is Clarke-subdifferentiable. Even though the cost function is subdifferentiable, only a limited number of algorithms for nonsmooth control problems in infinite dimensional spaces are available; see [21] and references therein. Our approach allows us to apply gradient-based solvers for PDE-constrained optimization problems. The ambiguity set (1.2) does not have a support constraint and, hence, it does not fulfill the assumptions made in [29]. The number of solutions of the parameterized state equation for the approach developed in [29] depends on the measure discretization. Our scheme is sampling-free and the evaluation of the smoothing function of (1.6) requires only a single solution of the state equation.

Taylor's expansions have also been used to approximate mean-plus-variance minimization problems [1, 12] and RO problems [2, 17, 27, 36]. The authors of [1, 12] have developed an approximation scheme for mean-plus-variance minimization problems with PDEs depending on a random field. They approximate the parameterized cost function $\widehat{J}(u, \cdot)$ using first- and second-order Taylor's expansions, allowing them to explicitly compute the expectation and variance of the surrogate objective function.

If the ambiguity set $\mathcal{P}$ is given by all probability distributions supported on a (compact) set $\Xi \subset \mathbb{R}^p$, that is, when $\mathcal{P} = \{ P \in \mathcal{M} : P(\Xi) = 1 \}$, the DROP (1.1) is equivalent to a RO problem [46, p. 535]. RO with PDEs has been considered in [2, 17, 27, 36, 22]. For numerical computations, the authors of [2, 27, 36] use second-order Taylor's expansions and obtain (1.5) without the value function defined by the SDP (1.7), that is, when $\varphi = 0$, and they either reformulate the TRP (1.8) using its necessary and sufficient optimality conditions, or use smooth optimization methods. The first approach results in a mathematical problem with complementarity problems [36, sect. 3.2.2]. Depending on the size of the trust-region radius Δ in (1.8), an optimal solution of (1.8) may be nonunique [39, p. 556] and, hence, the optimal value function (1.8) may be nondifferentiable. Optimization methods for smooth problems may therefore not be a suitable class of algorithms. Our algorithmic scheme provides an alternative to those used in [2, 27, 36] for RO with PDEs.

Outline. We describe our algorithmic approach for the solution of (1.5) in section 2. Section 3 provides the existence results of optimal solutions of the DROP (1.1), the approximated DROP (1.5) and the smoothed DROP (2.1) as well as existence of a worst-case distribution for the maximization problem in (1.1). We prove a convergence result for the homotopy method in section 4. The worst-case expected error of the reduced parameterized objective function (see (1.3)) and the quadratic approximation (see (1.4)) is examined in section 5. Section 6 is used to discuss the evaluation of the smoothing function of (2.1) and its derivative, and the complexity of our scheme in terms of PDE solutions. We provide applications and numerical results in section 7 and draw conclusions in section 8.

Notation. We define $\mathbb{R}_+ = [0, \infty)$ and $\mathbb{R}_{++} = (0, \infty)$. The set of symmetric $m \times m$-matrices and positive (semi)definite matrices is $\mathbb{S}^m$ and $\mathbb{S}^m_{++} \subset \mathbb{S}^m$ ($\mathbb{S}^m_+ \subset \mathbb{S}^m$), respectively, and the identity matrix is I. We use $\lambda : \mathbb{S}^p \to \mathbb{R}^p$ to denote the eigenvalue mapping with $\lambda_{\max}(A) = \lambda_1(A) \geq \cdots \geq \lambda_p(A) = \lambda_{\min}(A)$. For $A, B \in \mathbb{S}^m$, $A \preccurlyeq B$ abbreviates $B - A \in \mathbb{S}^m_+$. The Frobenius inner-product on $\mathbb{S}^m$ is $\bullet$, $\|A\|_2$ is the spectral norm of $A \in \mathbb{S}^p$, and $\| \cdot \|_B = \|B^{1/2} \cdot \|_2$ with $B \in \mathbb{S}^p_{++}$. For $a \in \mathbb{R}^m$, $\mathrm{Diag}(a)$ denotes the diagonal matrix with $(\mathrm{Diag}(a))_{ii} = a_i$ and $(\cdot)_+ = \max\{0, \cdot\}$. The Euclidean norm on $\mathbb{R}^m$ is $\| \cdot \|_2$ and $B_\epsilon(x) = \{ y \in \mathbb{R}^m : \|y - x\|_2 < \epsilon \}$ with $\epsilon > 0$. We denote the Fréchet derivative of $G : X_1 \times X_2 \to X_3$ w.r.t. x_1 by either $D_{x_1} G$ or G_{x_1}, and the gradient of G w.r.t. x_1 by $\nabla_{x_1} G$. If $A \in \mathscr{L}(X_1, X_2)$, then $A^* \in \mathscr{L}(X_2^*, X_1^*)$ is its adjoint operator. The set $\mathscr{L}(X_1, X_2)$ stands for the space of bounded linear operators from X_1 to X_2, $X^* = \mathscr{L}(X, \mathbb{R})$, and $\langle \cdot, \cdot \rangle_{X^*, X}$ is the duality pairing of X^* and X. The

Algorithm 2.1 Homotopy method

Choose $(\tau_1, \nu_1, \eta_1) > 0$ and $u_0 \in U_{\mathrm{ad}}$.
For $k = 1, 2, \dots$

1. Compute a stationary point u_k of (2.1) using u_{k-1} as initial point.
2. Choose $0 < (\tau_{k+1}, \nu_{k+1}, \eta_{k+1}) < (\tau_k, \nu_k, \eta_k)$.

map 1_A is the indicator function of A. Throughout, let $\mathcal{N}(\mu, \Sigma)$ denote the normal distribution with mean $\mu \in \mathbb{R}^p$ and covariance $\Sigma \in \mathbb{S}^p_+$. A metric space X is endowed with its Borel-σ-field. A function is measurable if it is Borel-measurable. We use ξ to denote a measurable mapping $\xi : \Omega \to \mathbb{R}^p$ and a deterministic vector $\xi \in \mathbb{R}^p$.

2. Algorithmic Approach to Solve the Approximated DROP.

We describe our algorithmic approach, which is based on [38], to compute a stationary point of the approximated DROP (1.5). We construct smoothing functions $\tilde{\psi} : U \times \mathbb{R}^2_{++} \to \mathbb{R}$ of $\psi : U \to \mathbb{R}$ (see (1.8)) and $\tilde{\varphi} : U \times \mathbb{R}_+ \to \mathbb{R}$ of $\varphi : U \to \mathbb{R}$ (see (1.7)), and compute approximate stationary points of a sequence of smoothed control problems

$$\min_{u \in U_{\mathrm{ad}}} \left\{ \widetilde{\Psi}(u; \tau_k, \nu_k, \eta_k) = \widehat{J}(u, \bar{\mu}) + \tilde{\varphi}(u; \tau_k) + \tilde{\psi}(u; \nu_k, \eta_k) \right\} \tag{2.1}$$

with decreasing smoothing parameters $(\tau_k, \nu_k, \eta_k) \in \mathbb{R}^3_{++}$ indexed by the outer iteration counter k, where $\widetilde{\Psi} : U \times \mathbb{R}^3_{++} \to \mathbb{R}$. We summarize this algorithmic framework in Algorithm 2.1. In section 4, we prove the convergence of weak limit points of optimal solutions, generated by Algorithm 2.1, to minimizers of (1.5). We can apply the same gradient-based optimization methods to (2.1) that are suitable for the nominal control problem

$$\min_{u \in U_{\mathrm{ad}}} \widehat{J}(u, \bar{\mu}). \tag{2.2}$$

We build our definition of a smoothing function on [13, Def. 3.1], allowing however for multiple smoothing parameters as in [38, Def. 3.1].

Definition 2.1. *Let X be a Banach space. The function $\tilde{\phi} : X \times \mathbb{R}^m_{++} \to \mathbb{R}$ is a* smoothing function *of $\phi : X \to \mathbb{R}$ if $\tilde{\phi}(\cdot; t)$ is continuously differentiable for every $t \in \mathbb{R}^m_{++}$, and for every $x \in X$, we have $|\phi(x) - \tilde{\phi}(x; t)| \le \gamma(t)$, where $\gamma : \mathbb{R}^m_+ \to \mathbb{R}_+$ fulfills $\gamma(t) \to 0$ as $\mathbb{R}^m_{++} \ni t \to 0$.*

In sections 2.1 and 2.2, we show that the smoothing functions for the optimal value functions (1.7) and (1.8) constructed in [38] satisfy Definition 2.1.

2.1. Smoothing Approach for the SDP.

We state a smoothing function of the optimal value function φ defined in (2.3), which was constructed in [38, sect. 4].

We require that $\nabla_{\xi\xi} \widehat{J}(\cdot, \bar{\mu})$ is continuously differentiable. From [53, Thm. 2.2], [38, sect. 4] and (1.7), we obtain that

$$\varphi(u) = (\bar{\sigma}_0/2) G(u) \bullet I + ((\bar{\sigma}_1 - \bar{\sigma}_0)/2) \sum_{i=1}^{p} \left(\lambda_i \left(G(u)\right)\right)_+, \tag{2.3}$$

where the preconditioned Hessian mapping $G : U \to \mathbb{S}^p$ is defined by

$$G(u) = \bar{\Sigma}^{1/2} \nabla_{\xi\xi} \widehat{J}(u, \bar{\mu}) \bar{\Sigma}^{1/2}. \tag{2.4}$$

We define $\tilde{\varphi} : U \times \mathbb{R}_{++} \to \mathbb{R}$ by

$$\tilde{\varphi}(u;\tau) = (\bar{\sigma}_0/2)G(u) \bullet I + ((\bar{\sigma}_1 - \bar{\sigma}_0)/2)\tilde{w}(\lambda(G(u));\tau), \tag{2.5}$$

where $\tilde{w} : U \times \mathbb{R}_{++} \to \mathbb{R}$ is given by

$$\tilde{w}(z;\tau) = \tau \sum_{i=1}^{p} \ln\left(1 + \exp\left(z_i/\tau\right)\right). \tag{2.6}$$

Now, let $\tau > 0$ be arbitrary. From [38, eq. (4.5)], we obtain

$$\varphi(u) \le \tilde{\varphi}(u;\tau) \le \varphi(u) + (1/2)\tau p \ln 2, \quad \text{for all} \quad u \in U. \tag{2.7}$$

The mapping $\mathbb{S}^p \ni A \mapsto \tilde{w}(\lambda(A);\tau)$ is twice continuously differentiable [37, Thm. 4.2]. Together with (2.7), we deduce that $\tilde{\varphi}$ is a smoothing function of φ. If $\nabla_{\xi\xi}\widehat{J}(\cdot,\bar{\mu})$ is twice continuously differentiable, then $\tilde{\varphi}(\cdot;\tau)$ is twice continuously differentiable.

2.2. Smoothing Approach for the TRP. We state a smoothing function of the optimal value function ψ (see (1.8)), which has been constructed in [38, sect. 5].

We require that $\nabla_\xi \widehat{J}(\cdot,\bar{\mu})$ and $\nabla_{\xi\xi}\widehat{J}(\cdot,\bar{\mu})$ are continuously differentiable. We define $\tilde{\psi} : U \times \mathbb{R}^2_{++} \to \mathbb{R}$ by

$$\tilde{\psi}(u;\nu,\eta) = \max_{\widetilde{s} \in \mathbb{R}^{p+2}} \left\{ \widetilde{g}_\nu(u)^T \widetilde{s} + (1/2)\widetilde{s}^T \widetilde{H}_\eta(u)\widetilde{s} : \ (1/2)\|\widetilde{s}\|_2^2 \le (1/2)\Delta^2 \right\} \tag{2.8}$$

where $\widetilde{H}_\eta : U \to \mathbb{S}^{p+2}$ and $\widetilde{g}_\nu : U \to \mathbb{R}^{p+2}$ are given by

$$\widetilde{H}_\eta(u) = \begin{bmatrix} G(u) & & \\ & 0 & \\ & & E(G(u);\eta) \end{bmatrix}, \qquad \widetilde{g}_\nu(u) = \begin{bmatrix} g(u) \\ \sqrt{2\nu} \\ \sqrt{2\nu} \end{bmatrix}, \tag{2.9}$$

and G is defined in (2.4), ν, $\eta > 0$, and the preconditioned gradient mapping $g : U \to \mathbb{R}^p$ is

$$g(u) = \bar{\Sigma}^{1/2} \nabla_\xi \widehat{J}(u,\bar{\mu}). \tag{2.10}$$

Here, $E : \mathbb{S}^p \times \mathbb{R}_{++} \to \mathbb{R}$ is the entropy function defined by

$$E(A;\eta) = \eta \ln \sum_{i=1}^{p} \exp(\lambda_i(A)/\eta). \tag{2.11}$$

We have $\lambda_{\max}(A) \le E(A;\eta) \le \lambda_{\max}(A) + \eta \ln p$ [40, eq. (17)], and $E(\cdot;\eta)$ is twice continuously differentiable for all $\eta > 0$ [37, Thm. 4.2]. The TRP (2.8) is a dual of a reciprocal barrier approximation of a smoothed dual of the TRP (1.8) [38, sect. 5.2].

Let ν, $\eta > 0$ be arbitrary. From [38, eq. (5.31)], we obtain that

$$\psi(u) \le \tilde{\psi}(u;\nu,\eta) \le \psi(u) + 2\sqrt{2\nu}\Delta + (1/2)\Delta^2 \eta \ln p \quad \text{for all} \quad u \in U. \tag{2.12}$$

Using the reasoning as in the proof of [38, Thm. 5.9], we can show that $\tilde{\psi}(\cdot;\nu,\eta)$ is continuously differentiable. Combined with (2.12), we obtain that it is a smoothing function of ψ. If $\nabla_\xi \widehat{J}(\cdot,\bar{\mu})$ and $\nabla_{\xi\xi}\widehat{J}(\cdot,\bar{\mu})$ are twice continuously differentiable, we can show that $\tilde{\psi}(\cdot;\nu,\eta)$ is twice continuously differentiable as in [38, Thm. 5.9].

3. Existence of Optimal Solutions. We prove the existence of optimal solutions of the DROPs (1.1), (1.5) and (2.1), and of a worst-case distribution of the maximization problem in (1.1).

3.1. Existence of Optimal Solutions to the DROP. We state conditions implying that (1.1) has optimal solutions which are built on those used in [32, Chap. 3] and [34, sect. 3.2].

Assumption 3.1. For all $(u, \xi) \in U_{\mathrm{ad}} \times \mathbb{R}^p$, the state equation $e(y, u, \xi) = 0$ has a unique solution $y = S(u, \xi) \in Y$, which defines the solution operator $S : U_{\mathrm{ad}} \times \mathbb{R}^p \to Y$.
1. The operator $S(\cdot, \xi) : U_{\mathrm{ad}} \to Y$ is weakly-weakly continuous for all $\xi \in \mathbb{R}^p$.
2. The mapping $S(u, \cdot) : \mathbb{R}^p \to Y$ is continuous for all $u \in U_{\mathrm{ad}}$.

Assumption 3.1 implies that the reduced parameterized objective function $\widehat{J}$ (see (1.3)) is well-defined. If Assumption 3.1 holds and Y is separable, then $S(u, \cdot) : \mathbb{R}^p \to Y$ is strongly measurable (see [26, Lem. 1.5] and [23, Cor. 2, p. 73]) and, hence, [34, Assumption 2.1.1] is satisfied, that is, strong measurability of $S(u, \cdot)$ for all $u \in U_{\mathrm{ad}}$. Assumption 3.1.1 may be verified using Lemma B.1. Assumption 3.1.2 can be shown using the implicit function theorem when $e_y(S(u, \xi), u, \xi) \in \mathscr{L}(Y, Z)$ is boundedly invertible for all $(u, \xi) \in U_{\mathrm{ad}} \times \mathbb{R}^p$.

Assumption 3.2. The function $J : Y \times U_{\mathrm{ad}} \times \mathbb{R}^p \to \mathbb{R}$ is continuous.
1. There exists $\rho \in \mathbb{R}$ such that $J(y, u, \xi) \geq \rho$ for all $y \in Y$, $u \in U_{\mathrm{ad}}$ and every $\xi \in \mathbb{R}^p$.
2. The mapping $J(\cdot, \cdot, \xi) : Y \times U_{\mathrm{ad}} \to \mathbb{R}$ is weakly lower semicontinuous for all $\xi \in \mathbb{R}^p$.

Assumption 3.2 is satisfied for, e.g., tracking type functionals; see Example 3.5. If $J(\cdot, \cdot, \xi)$ is convex for all $\xi \in \mathbb{R}^p$ and J is continuous, then Assumption 3.2.3 holds [24, Thm. 1.18].

We introduce the notion of uniform integrability. A measurable function $\theta : \mathbb{R}^p \to \mathbb{R}$ is uniformly integrable (w.r.t. $\mathcal{P}$) if $\sup_{P \in \mathcal{P}} \mathbb{E}_P[|\theta(\xi)| 1_{|\theta(\xi)| \geq t}] \to 0$ as $t \to \infty$ [8, sect. 2.7(i)].

Assumption 3.3. For all $u \in U_{\mathrm{ad}}$, $\widehat{J}(u, \cdot)$ (see (1.3)) is uniformly integrable.

Assumption 3.3 can be verified for many problems using the next lemma.

LEMMA 3.4. *The following conditions imply that the measurable function* $h : \mathbb{R}^p \to \mathbb{R}$ *is uniformly integrable: (i)* $\sup_{P \in \mathcal{P}} \mathbb{E}_P[|h(\xi)|^r] < \infty$ *for some* $r > 1$*; (ii) there exists a uniformly integrable function* $g : \mathbb{R}^p \to \mathbb{R}_+$ *such that* $|h| \leq g$*; (iii)* g_1, $g_2 : \mathbb{R}^p \to \mathbb{R}$ *are uniformly integrable and* $h = g_1 + g_2$.

Proof. Fix $t > 0$ and $P \in \mathcal{P}$. The following estimates yield the claim: (i) $\mathbb{E}_P[|h(\xi)| 1_{|h(\xi)| > t}] \leq t^{-r+1} \mathbb{E}_P[|h(\xi)|^r]$ (see, e.g., [26, p. 44]); (iii) $\mathbb{E}_P[|h(\xi)| 1_{|h(\xi)| \geq 2t}] \leq 2\mathbb{E}_P[|g_1(\xi)| 1_{|g_1(\xi)| \geq t}] + 2\mathbb{E}_P[|g_2(\xi)| 1_{|g_2(\xi)| \geq t}]$ (see, e.g., [7, p. 230]). ∎

Example 3.5. Let X be a Hilbert space, $Y \hookrightarrow X$ be a continuous embedding, and let Assumption 3.1 hold. Consider $J : Y \times U \to \mathbb{R}_+$ defined by $J(y, u) = (1/2)\|y - y_d\|_X^2 + (\alpha/2)\|u\|_U^2$, where $y_d \in X$ and $\alpha \geq 0$. Let $u \in U$ be arbitrary. If $\|S(u, \cdot)\|_X^2$ is uniformly integrable, then $\widehat{J}(u, \cdot)$ is uniformly integrable. Indeed, Young's inequality yields

$$\widehat{J}(u, \xi) = \tfrac{1}{2}\|S(u, \xi) - y_d\|_X^2 + \tfrac{\alpha}{2}\|u\|_U^2 \leq \|S(u, \xi)\|_X^2 + \|y_d\|_X^2 + \tfrac{\alpha}{2}\|u\|_U^2, \tag{3.1}$$

for all $\xi \in \mathbb{R}^p$. Now, Lemma 3.4 implies the assertion. We verify the uniform integrability of $\|S(u, \cdot)\|_X^2$ for two examples in section 7.

We show that the DROP (1.1) has an optimal solution.

THEOREM 3.6. *Let Assumptions 3.1–3.3 hold. Suppose the level set* $\{ u \in U_{ad} : f(u) \leq \gamma \}$ *is nonempty and bounded for some* $\gamma \in \mathbb{R}$*, and* $U_{ad} \subset U$ *is closed and convex, where* f *is defined in* (1.9)*. Then, the DROP* (1.1) *has an optimal solution.*

We apply Lemmas 3.7 and 3.8 to prove Theorem 3.6.

LEMMA 3.7. *We have* $\{ \mathcal{N}(\mu, \Sigma) : \|\bar{\Sigma}^{-1/2}(\mu - \bar{\mu})\|_2 \leq \Delta,\ \sigma_0\bar{\Sigma} \preccurlyeq \Sigma \preccurlyeq \sigma_1\bar{\Sigma} \} \subset \mathcal{P}$.

Proof. The claim follows from a discussion in [10, pp. 185–186]. ∎

LEMMA 3.8. *If Assumptions 3.1–3.3 hold, then* $f : U_{ad} \to \mathbb{R}$ *defined in* (1.9) *is finite-valued and weakly lower semicontinuous.*

Proof. Fix $u \in U_{\mathrm{ad}}$ and $\delta > 0$. Since $\mathcal{P} \neq \emptyset$ (see Lemma 3.7) and Assumption 3.3 holds, for some $t > 0$ and all $P \in \mathcal{P}$, we get $\mathbb{E}_P[|\widehat{J}(u,\xi)|] \leq t + \mathbb{E}_P[|\widehat{J}(u,\xi)| 1_{|\widehat{J}(u,\xi)| \geq t}] \leq t + \delta$. Hence $f(u) \in \mathbb{R}$. Since $u \in U_{\mathrm{ad}}$ is arbitrary, f is finite-valued.

Now, fix $(u_k) \subset U_{\mathrm{ad}}$ with $u_k \rightharpoonup u \in U_{\mathrm{ad}}$ as $k \to \infty$, and $P \in \mathcal{P}$. We show that

$$\liminf_{k\to\infty} \mathbb{E}_P[\widehat{J}(u_k,\xi)] \geq \mathbb{E}_P[\widehat{J}(u,\xi)]. \tag{3.2}$$

Assumptions 3.1 and 3.2 imply that $\widehat{J}(\cdot,\xi)$ is weakly lower semicontinuous for all $\xi \in \mathbb{R}^p$ and that $\widehat{J}(u,\cdot)$ is continuous and, hence, measurable for all $u \in U_{\mathrm{ad}}$ [26, Lem. 1.5]. We deduce $\mathbb{E}_P[\liminf_{k\to\infty} \widehat{J}(u_k,\xi)] \geq \mathbb{E}_P[\widehat{J}(u,\xi)]$. Fatou's lemma yields (3.2).

It must yet be shown that f (see (1.9)) is weakly lower semicontinuous. Fix $\varepsilon > 0$. Since $f(u) < \infty$, there exists $P_\varepsilon \in \mathcal{P}$ with $f(u) \leq \mathbb{E}_{P_\varepsilon}[\widehat{J}(u,\xi)] + \varepsilon$. Now, (3.2) yields

$$f(u) \leq \mathbb{E}_{P_\varepsilon}[\widehat{J}(u,\xi)] + \varepsilon \leq \liminf_{k\to\infty}\ \mathbb{E}_{P_\varepsilon}[\widehat{J}(u_k,\xi)] + \varepsilon \leq \liminf_{k\to\infty}\ f(u_k) + \varepsilon.$$

Since $\varepsilon > 0$, $(u_k) \subset U_{\mathrm{ad}}$ and $u \in U_{\mathrm{ad}}$ are arbitrary, f is weakly lower semicontinuous. ∎

Proof of Theorem 3.6. Owing to Lemma 3.8, we can apply the direct method of the calculus of variations to prove that (1.1) has an optimal solution $u^* \in U_{\mathrm{ad}}$. ∎

3.2. Existence of Worst-Case Distributions. We show that a worst-case distribution of the maximization problem in (1.1) exists. A worst-case distribution of (1.5) is the normal distribution, where the mean is a maximizer of (1.8) and the covariance matrix is one of (1.7).

THEOREM 3.9. *If Assumptions 3.1–3.3 hold and* $u \in U_{ad}$*, then there exists* $P^* \in \mathcal{P}$ *such that* $\mathbb{E}_{P^*}[\widehat{J}(u,\xi)] = \sup_{P\in\mathcal{P}} \mathbb{E}_P[\widehat{J}(u,\xi)]$.

We use Lemmas 3.10 and 3.11 to prove Theorem 3.9.

LEMMA 3.10. *If* $(P_k) \subset \mathcal{P}$ *fulfills* $P_k \rightharpoonup P$ *as* $k \to \infty$ *with* $P \in \mathcal{M}$*, then* $P \in \mathcal{P}$.[1]

Proof. Let i, $j \in \{1,\ldots,p\}$ and $d \in \mathbb{R}^p$ be arbitrary. Define θ_1, θ_2, $\theta_3 : \mathbb{R}^p \to \mathbb{R}$ by $\theta_1(\xi) = \xi_i$, $\theta_2(\xi) = \xi_i\xi_j$ and $\theta_3(\xi) = \exp(d^T\xi)$. We show that they are uniformly integrable. We have $|\theta_1(\xi)| \leq \|\xi\|_2$ and $|\theta_2(\xi)| \leq \|\xi\|_2^2$ for all $\xi \in \mathbb{R}^p$. Lemma A.1 and (1.2) imply $\sup_{P\in\mathcal{P}} \mathbb{E}_P[\|\xi\|_2^r] < \infty$ for $r = 1, 2, 4$. For all $P \in \mathcal{P}$, we have

$$\mathbb{E}_{P_k}[|\theta_3(\xi)|^2] = \exp(2d^T\mathbb{E}_P[\xi])\mathbb{E}_P[\exp(2d^T(\xi - \mathbb{E}_P[\xi]))] \leq \exp(2d^T\mathbb{E}_P[\xi] + 2\sigma_1 d^T\bar{\Sigma}d).$$

Lemma 3.4 implies that θ_1, θ_2 and θ_3 are uniformly integrable. Now, [8, Thm. 2.7.1] yields $\mathbb{E}_{P_k}[\xi_i] = \mathbb{E}_{P_k}[\theta_1(\xi)] \to \mathbb{E}_P[\theta_1(\xi)] = \mathbb{E}_P[\xi_i]$, and $\mathbb{E}_{P_k}[\xi_i\xi_j] = \mathbb{E}_{P_k}[\theta_2(\xi)] \to \mathbb{E}_P[\theta_2(\xi)] = \mathbb{E}_P[\xi_i\xi_j]$ as $k \to \infty$. Since i, $j \in \{1,\ldots,p\}$ are arbitrary, we obtain

$$\mathbb{E}_{P_k}[\xi] \to \mathbb{E}_P[\xi], \quad \text{and} \quad \mathbb{E}_{P_k}[\xi\xi^T] \to \mathbb{E}_P[\xi\xi^T] \quad \text{as} \quad k \to \infty.$$

[1] We refer to [26, p. 39] for a definition of weak convergence of a sequence of probability measures.

Combined with [8, Thm. 2.7.1], we get

$$\begin{aligned}\mathrm{Cov}_{P_k}[\xi] = \mathbb{E}_{P_k}[\xi\xi^T] - \mathbb{E}_{P_k}[\xi]\mathbb{E}_{P_k}[\xi]^T &\to \mathrm{Cov}_P[\xi],\\ \mathbb{E}_{P_k}[\exp(d^T(\xi - \mathbb{E}_{P_k}[\xi]))] = \exp(-d^T\mathbb{E}_{P_k}[\xi])\mathbb{E}_{P_k}[\theta_3(\xi)] &\to \mathbb{E}_P[\exp(d^T(\xi - \mathbb{E}_P[\xi]))].\end{aligned}$$

We have $\|\bar{\Sigma}^{-1/2}(\mathbb{E}_{P_k}[\xi] - \bar{\mu})\|_2 \leq \Delta$, $\sigma_0\bar{\Sigma} \preccurlyeq \mathrm{Cov}_{P_k}[\xi] \preccurlyeq \sigma_1\bar{\Sigma}$, and $\mathbb{E}_{P_k}[\exp(d^T(\xi - \mathbb{E}_{P_k}[\xi]))] \leq \exp((\sigma_1/2)d^T\bar{\Sigma}d)$ for all $k \in \mathbb{N}_0$ and $d \in \mathbb{R}^p$. Putting together the pieces, we conclude that $\mathcal{P} \in \mathcal{P}$. ◻

LEMMA 3.11. *If $(P_k) \subset \mathcal{P}$ is a sequence of probability measures, then (P_k) has a weakly convergent subsequence $(P_k)_K$ such that $P_k \rightharpoonup P \in \mathcal{M}$ as $K \ni k \to \infty$.*

Proof. We show that (P_k) is tight.[2] Lemma A.1 yields $\sup_{k\in\mathbb{N}_0} \mathbb{E}_{P_k}[\|\xi\|_2^2] < \infty$. Markov's inequality gives $\sup_{k\in\mathbb{N}_0} P_k(\|\xi\|_2 > \sqrt{t}) \leq (\sup_{k\in\mathbb{N}_0} \mathbb{E}_{P_k}[\|\xi\|_2^2])/t \to 0$ as $t \to \infty$ and, hence, (P_k) is tight. Combined with [26, Lem. 4.20 and Prop. 4.21], we conclude that (P_k) has a subsequence $(P_k)_K$ such that $P_k \rightharpoonup P \in \mathcal{M}$ as $K \ni k \to \infty$. ◻

Proof of Theorem 3.9. Lemma 3.8 yields $\sup_{P\in\mathcal{P}} \mathbb{E}_P[\widehat{J}(u,\xi)] \in \mathbb{R}$. Let $(P_k) \subset \mathcal{P}$ satisfy $\lim_{k\to\infty} \mathbb{E}_{P_k}[\widehat{J}(u,\xi)] = \sup_{P\in\mathcal{P}} \mathbb{E}_P[\widehat{J}(u,\xi)]$. Lemma 3.11 implies the existence of a subsequence $(P_k)_K$ of (P_k) such that $P_k \rightharpoonup P^* \in \mathcal{M}$ as $K \ni k \to \infty$. Lemma 3.10 yields $P^* \in \mathcal{P}$. Assumptions 3.1–3.3 imply that $\widehat{J}(u,\cdot)$ is continuous and uniformly integrable. Hence, [8, Thm. 2.7.1] yields $\mathbb{E}_{P^*}[\widehat{J}(u,\xi)] = \lim_{K\ni k\to\infty} \mathbb{E}_{P_k}[\widehat{J}(u,\xi)] = \sup_{P\in\mathcal{P}} \mathbb{E}_P[\widehat{J}(u,\xi)]$. ◻

3.3. Existence of Optimal Solutions to Approximated and Smoothed DROPs. We show that the approximated DROP (1.5), and the smoothed DROP (2.1) have optimal solutions.

Assumption 3.12. For some $\epsilon > 0$ and all $(u,\xi) \in U_{\mathrm{ad}} \times B_\epsilon(\bar{\mu})$, $e(y,u,\xi) = 0$ has a unique solution $y = S(u,\xi) \in Y$, which defines the operator $S : U_{\mathrm{ad}} \times B_\epsilon(\bar{\mu}) \to Y$.
1. For all $u \in U_{\mathrm{ad}}$, $J(\cdot,u,\cdot)$ is twice continuously differentiable, where $J : Y \times U_{\mathrm{ad}} \times B_\epsilon(\bar{\mu}) \to \mathbb{R}$.
2. The map $e : Y \times U \times B_\epsilon(\bar{\mu}) \to Z$ is twice continuously differentiable. For all $(u,\xi) \in U_{\mathrm{ad}} \times B_\epsilon(\bar{\mu})$, the operator $e_y(S(u,\xi),u,\xi) \in \mathscr{L}(Y,Z)$ is boundedly invertible.
3. The mapping $\widehat{J}(\cdot,\bar{\mu}) : U_{\mathrm{ad}} \to \mathbb{R}$ is weakly lower semicontinuous, and $\nabla_\xi\widehat{J}(\cdot,\bar{\mu}) : U_{\mathrm{ad}} \to \mathbb{R}^p$ and $\nabla_{\xi\xi}\widehat{J}(\cdot,\bar{\mu}) : U_{\mathrm{ad}} \to \mathbb{S}^p$ are weakly(-strongly) continuous.

Assumption 3.12.1–2 imply that the objective functions of the approximated DROP (1.5) and the smoothed DROP (2.1) are well-defined. We comment on Assumption 3.12.3 in Remark 3.15. Assumption 3.12 implies that both DROPs (1.5) and (2.1) have optimal solutions.

THEOREM 3.13. *Let Assumption 3.12 hold, $U_{ad} \subset U$ be nonempty, closed and convex. Furthermore, let U_{ad} be bounded or Ψ be coercive, where Ψ is defined in (1.6). Then, the approximated DROP (1.5) has an optimal solution and, for every $\rho^k = (\tau_k,\nu_k,\eta_k) \in \mathbb{R}^3_{++}$, the smoothed DROP (2.1) has an optimal solution.*

We prove Theorem 3.13 using Lemma 3.14.

LEMMA 3.14. *Let Assumption 3.12 hold and $\rho^k = (\tau_k,\nu_k,\eta_k) \in \mathbb{R}^3_{++}$ be arbitrary. Then, $\Psi : U_{ad} \to \mathbb{R}$ and $\widetilde{\Psi}(\cdot;\rho^k) : U_{ad} \to \mathbb{R}$ are weakly lower semicontinuous.*

Proof. The mapping λ is continuous [25, Cor. 6.3.8] and, hence, $\lambda \circ G : U_{\mathrm{ad}} \to \mathbb{R}$ and $E(\cdot;\eta) \circ G : U_{\mathrm{ad}} \to \mathbb{R}$ are weakly continuous, where G is defined in (2.4) and

[2]We refer to [26, p. 62] for a definition of tightness of a sequence of probability measures.

$E(\cdot;\eta_k)$ in (2.11). From (2.3) and (2.5), we obtain that $\varphi : U_{\text{ad}} \to \mathbb{R}$ and $\tilde{\varphi}(\cdot;\tau_k) : U_{\text{ad}} \to \mathbb{R}$ are weakly continuous. Owing to Assumption 3.12, [22, Thm. 2.5] implies that $\psi : U_{\text{ad}} \to \mathbb{R}$ (see (1.8)) and $\widetilde{\psi}(\cdot;\nu_k,\eta_k) : U_{\text{ad}} \to \mathbb{R}$ (see (2.8)) are weakly lower semicontinuous. The weak lower semicontinuity of $\widehat{J}(\cdot,\bar{\mu}) : U_{\text{ad}} \to \mathbb{R}$ implies that of $\Psi : U_{\text{ad}} \to \mathbb{R}$ and $\widetilde{\Psi}(\cdot;\rho^k) : U_{\text{ad}} \to \mathbb{R}$. ◻

Proof of Theorem 3.13. Lemma 3.14 yields the lower semicontinuity of Ψ and $\widetilde{\Psi}(\cdot;\rho^k)$. If Ψ is coercive, then (2.7) and (2.12) imply that $\widetilde{\Psi}(\cdot;\rho^k)$ is coercive. Now, the direct method of the calculus of variations yields the existence of an optimal solution of (1.5) and of (2.1). ◻

Assumption 3.12.3 may be verified using compact embeddings.

Remark 3.15. Let $U_{\text{ad}} \subset U$ and $U \hookrightarrow U_0$ be a compact embedding, and $\wp : U \to \mathbb{R}$ be weakly lower semicontinuous. We consider $J_1 : Y \times B_\epsilon(\bar{\mu}) \to \mathbb{R}$ and the solution operator $S : U_0 \times B_\epsilon(\bar{\mu}) \to Y$. Let J_1 and S be twice continuously differentiable. Suppose that $J(y,u,\xi) = J_1(y,\xi) + \wp(u)$ for all $y \in Y$, $u \in U$ and $\xi \in B_\epsilon(\bar{\mu})$. Then Assumption 3.12.1 and 3.12.3 hold. Indeed, $\widehat{J}(\cdot,\bar{\mu}) : U_{\text{ad}} \to \mathbb{R}$ is weakly lower semicontinuous. Moreover, $U_0 \times \mathbb{R}^p \ni (u,\xi) \mapsto J_1(S(u,\xi);\xi)$ is twice continuously differentiable, and we have $D_\xi \widehat{J} = D_\xi(J(S(\cdot,\cdot),\cdot))$ and $D_{\xi\xi}\widehat{J} = D_{\xi\xi}(J(S(\cdot,\cdot),\cdot))$. We deduce that $\nabla_\xi \widehat{J}(\cdot,\bar{\mu}) : U_{\text{ad}} \to \mathbb{R}^p$ and $\nabla_{\xi\xi}\widehat{J}(\cdot,\bar{\mu}) : U_{\text{ad}} \to \mathbb{S}^p$ are weakly continuous.

4. Global Convergence of the Homotopy Method. We present a convergence result for a sequence of optimal solutions of (2.1) as the smoothing parameters converge to zero. Theorem 4.1 implies global convergence of a sequence of minimizers generated by Algorithm 2.1.

THEOREM 4.1. *Let the conditions of Theorem 3.13 hold, $\rho^k = (\tau_k,\nu_k,\eta_k) \in \mathbb{R}^3_{++}$ fulfill $\rho^k \to 0$ as $k \to \infty$ and, for all $k \geq 0$, let u_k be an optimal solution of (2.1). Then (u_k) is bounded and each weak limit point of (u_k) is an optimal solution of (1.5).*

Proof. Fix $u \in U_{\text{ad}}$. Using (2.7) and (2.12), and $\widetilde{\Psi}(u_k;\rho^k) \leq \widetilde{\Psi}(u;\rho^k)$, we have

$$\Psi(u_k) \leq \widetilde{\Psi}(u_k;\rho^k) \leq \widetilde{\Psi}(u;\rho^k) \leq \Psi(u) + \tfrac{1}{2}\tau_k p \ln 2 + 2\sqrt{2\nu_k}\Delta + (1/2)\Delta^2\eta_k \ln p, \tag{4.1}$$

for all $k \in \mathbb{N}_0$, where Ψ is defined in (1.6) and $\widetilde{\Psi}$ in (2.1). The sequence $(u_k) \subset U_{\text{ad}}$ is bounded because either U_{ad} is bounded or Ψ is coercive by assumption.

Let u^* be a weak accumulation point of (u_k). Then, there exists $(u_k)_K \subset (u_k)$ such that $u_k \rightharpoonup u^*$ as $K \ni k \to \infty$. We have $u^* \in U_{\text{ad}}$ since U_{ad} is closed and convex. Lemma 3.14 yields $\Psi(u^*) \leq \liminf_{K\ni k\to\infty} \Psi(u_k)$. Combined with (4.1) and $\rho_k \to 0$ as $k \to \infty$, we obtain $\Psi(u^*) \leq \Psi(u)$. Hence, u^* is an optimal solution of (1.5). ◻

5. Error of Quadratic Approximation. We show that the worst-case expected error between the cost function of (1.1) and that of (1.5) converges to zero for "shrinking" ambiguity sets.

LEMMA 5.1. *Let $u \in U_{ad}$ be arbitrary, Assumption 3.12.1-2 hold, and $L(u,\cdot) : \mathbb{R}^p \to \mathbb{R}_+$ be measurable such that $\sup_{P\in\mathcal{P}} \mathbb{E}_P[|L(u,\xi)|^2] < \infty$ and*

$$|\widehat{J}(u,\xi) - Q(u,\xi;\bar{\mu})| \leq (L(u,\xi)/6)\|\xi - \bar{\mu}\|_2^3, \quad \textit{for all} \quad \xi \in \mathbb{R}^p, \tag{5.1}$$

where $\widehat{J}$ is defined in (1.3) and Q in (1.4). Then, we have

$$\sup_{P\in\mathcal{P}} \mathbb{E}_P[|\widehat{J}(u,\xi) - Q(u,\xi;\bar{\mu})|] \to 0 \quad \text{as} \quad (\Delta,\sigma_1) \to 0^+. \tag{5.2}$$

Proof. Let $P \in \mathcal{P}$ be arbitrary. Using Hölder's inequality and (5.1), we find that

$$\mathbb{E}_P[|\widehat{J}(u,\xi) - Q(u,\xi;\bar{\mu})|] \le (1/6)(\mathbb{E}_P[|L(u,\xi)|^2])^{1/2}(\mathbb{E}_P[\|\xi - \bar{\mu}\|_2^6])^{1/2}. \tag{5.3}$$

The triangle inequality, and the monotonicity and convexity of $\mathbb{R}_+ \ni z \mapsto z^6$ imply

$$\mathbb{E}_P[\|\xi - \bar{\mu}\|_2^6] \le 2^5\mathbb{E}_P\left[\|\xi - \mathbb{E}_P[\xi]\|_2^6\right] + 2^5\|\mathbb{E}_P[\xi] - \bar{\mu}\|_2^6.$$

Lemma A.1 yields $\mathbb{E}_P[\|\xi - \mathbb{E}_P[\xi]\|_2^6] \le 2(6/\mathrm{e})^6(I \bullet \sigma_1\bar{\Sigma})^6$. Using (1.2), we obtain $\|\bar{\Sigma}^{-1/2}(\mathbb{E}_P[\xi] - \bar{\mu})\|_2 \le \Delta$ and, hence, $\|\mathbb{E}_P[\xi] - \bar{\mu}\|_2^6 \le \|\bar{\Sigma}^{1/2}\|_2^6\Delta^6$. We deduce

$$\mathbb{E}_P[\|\xi - \bar{\mu}\|_2^6] \le 64(6/\mathrm{e})^6(I \bullet \sigma_1\bar{\Sigma})^6 + 32\|\bar{\Sigma}^{1/2}\|_2^6\Delta^6.$$

Hence $\sup_{P\in\mathcal{P}} \mathbb{E}_P[\|\xi - \bar{\mu}\|_2^6] \to 0$ as $(\Delta, \sigma_1) \to 0$. Combined with (5.3), we get (5.2). ∎

6. Quadratic Approximation of the Reduced Parametrized Objective Function. We compute the derivatives of the smoothing functions $\tilde{\varphi}$ (see (2.5)) and $\tilde{\psi}$ (see (2.8)). Furthermore, we discuss the complexity of computing the gradient and the Hessian which define the quadratic model Q (see (1.4)), and of the first derivative of the smoothing functions $\tilde{\varphi}$ and $\tilde{\psi}$. Throughout the section, let Assumption 3.12.1–2 hold, $J(y,\cdot,\xi) : U \to \mathbb{R}$ be continuously differentiable for all $(y,\xi) \in Y \times \mathbb{R}^p$, and let $(\tau,\nu,\eta) \in \mathbb{R}^3_{++}$ and $\bar{u} \in U$ be arbitrary.

6.1. Evaluation of the Optimal Value Functions and their Derivatives. We derive formulas for the derivative of the smoothing functions $\tilde{\varphi}$ (see (2.5)) and $\tilde{\psi}$ (see (2.8)).

6.1.1. Evaluation of the Smoothing Function of the SDP and its Derivative. In order to evaluate $\tilde{\varphi}$, we propose computing the matrix $\nabla_{\xi\xi}J(\bar{u},\bar{\mu})$ when the number of parameters p is moderate. We use the identity $G(\bar{u}) \bullet I = \nabla_{\xi\xi}J(\bar{u},\bar{\mu}) \bullet \bar{\Sigma}$ to evaluate the first addend in (2.5) and compute an eigendecomposition of $G(\bar{u})$ via the (generalized) eigenvalue problem $\nabla_{\xi\xi}J(u,\bar{\mu})q = \lambda\bar{\Sigma}^{-1}q$, $\|q\|_2 = 1$, where G is defined in (2.4). We have

$$\langle DG(\bar{u})^*P, s\rangle_{U^*,U} = P \bullet (DG(\bar{u})s) = \langle D(G(\bar{u}) \bullet P), s\rangle_{U^*,U} \tag{6.1}$$

for all $s \in U$, $P \in \mathbb{S}^p$. From section 2.1, [37, Lem. 3.1], (2.5) and (6.1), we obtain that

$$D_u\tilde{\varphi}(\bar{u};\tau) = (\bar{\sigma}_0/2)DG(\bar{u})^*I + ((\bar{\sigma}_1 - \bar{\sigma}_0)/2)DG(\bar{u})^*[Q(\bar{u})M(\bar{u})Q(\bar{u})^T], \tag{6.2}$$

where $\tilde{w}$ is defined in (2.6) and G in (2.4), $Q(\bar{u}) \in \mathbb{R}^{p\times p}$ fulfills $Q(\bar{u})^TQ(\bar{u}) = I$, and

$$G(\bar{u}) = Q(\bar{u})\mathrm{Diag}(\lambda(G(\bar{u})))Q(\bar{u})^T, \quad \text{and} \quad M(\bar{u}) = \mathrm{Diag}(\nabla_x\tilde{w}(\lambda(G(\bar{u}));\tau)). \tag{6.3}$$

Using (6.1), the expression $DG(\bar{u})^*[Q(\bar{u})M(\bar{u})Q(\bar{u})^T]$ in (6.2) becomes

$$DG(\bar{u})^*[Q(\bar{u})M(\bar{u})Q(\bar{u})^T] = \sum_{i=1}^{p} m_{ii}(\bar{u})\ D_u(q_i(\bar{u})^TG(u)q_i(\bar{u}))\big|_{u=\bar{u}}, \tag{6.4}$$

where $m_{ii}(\bar{u})$ is the ith diagonal entry of $M(\bar{u})$ and $q_i(\bar{u})$ the ith column of $Q(\bar{u})$.

6.1.2. Evaluation of the Smoothing Function of the TRP and its Derivative. In order to evaluate $\tilde{\psi}$, we compute $E(G(\bar{u});\eta)$ (see (2.11)) using the eigenvalues of $G(\bar{u})$ that are already needed to evaluate $\tilde{\varphi}$; see section 6.1.1. From section 2.2 and [9, Rem. 4.14], we obtain that

$$D_u\tilde{\psi}(\bar{u};\nu,\eta) = D_u(g(\bar{u})^T s^*) + \tfrac{1}{2}D_u((s^*)^T G(\bar{u})s^*) + \tfrac{1}{2}\tilde{s}_{p+2}^2 D_u E(G(\bar{u});\eta), \tag{6.5}$$

where $\widetilde{s} = (s^*, \tilde{s}_{p+1}, \tilde{s}_{p+2}) \in \mathbb{R}^{p+2}$ is the optimal solution of (2.8) for $u = \bar{u}$. We have

$$\nabla_A E(A;\eta) = R(A)\mathrm{Diag}(\theta(A))R(A)^T \quad \text{and} \quad \theta_i(A;\eta) = \frac{\exp(\lambda_i(A)/\eta)}{\sum_{i=1}^p \exp(\lambda_i(A)/\eta)}, \tag{6.6}$$

where $R(A) \in \mathbb{R}^{p\times p}$, $R(A)^T R(A) = I$, and $A = R(A)\mathrm{Diag}(\lambda(A))R(A)^T \in \mathbb{S}^p$ [40, eq. (18)]. Combined with the chain rule and (6.1), we find that

$$D_u E(G(\bar{u});\eta) = \sum_{i=1}^p \theta_i(\bar{u})\ D_u(q_i(\bar{u})^T G(u) q_i(\bar{u}))\big|_{u=\bar{u}}\,, \tag{6.7}$$

where $\theta_i(\bar{u}) = \theta_i(G(\bar{u});\eta)$ and $q_i(\bar{u})$ is the ith column of $Q(\bar{u})$ (see (6.3)).

6.2. Computation of Derivatives and Computational Complexity. We discuss the number of linear and nonlinear PDE solves required to evaluate $\widetilde{\Psi}(\cdot;\tau,\nu,\eta)$ (see (2.1)) and its derivative. We define $\rho = (\tau,\nu,\eta)$.

To compute $\widetilde{\Psi}(\bar{u};\rho)$, we evaluate $\widehat{J}(\bar{u},\bar{\mu})$, $\nabla_\xi \widehat{J}(\bar{u},\bar{\mu})$ and $\nabla_{\xi\xi}\widehat{J}(\bar{u},\bar{\mu})$. We evaluate $\widehat{J}(\bar{u},\bar{\mu})$ using the solution of the state equation. The gradient $\nabla_\xi \widehat{J}(\bar{u},\bar{\mu})$ can be computed via the adjoint approach which requires the solution of an adjoint equation [24, eq. (1.85)]. We compute the Hessian $\nabla_{\xi\xi}\widehat{J}(\bar{u},\bar{\mu})$ via the adjoint approach which requires the solution of p sensitivity and p adjoint equations [24, sect. 1.6.5].

To evaluate the derivate of $\widetilde{\Psi}(\cdot;\rho)$ at $\bar{u}$, we compute $D_u\widehat{J}(\bar{u},\bar{\mu})$, $D_u(\nabla_\xi \widehat{J}(\bar{u},\bar{\mu})^T s_\xi)$, $D_u(s_\xi^T \nabla_{\xi\xi}\widehat{J}(\bar{u},\bar{\mu})s_\xi)$, and $D_u(q_i(\bar{u})^T\nabla_{\xi\xi}\widehat{J}(\bar{u},\bar{\mu})q_i(\bar{u}))$ for $i = 1,\ldots,p$, and a certain $s_\xi \in \mathbb{R}^p$; see (6.4), (6.5) and (6.7). We can evaluate $D_u\widehat{J}(\bar{u},\bar{\mu})$ without further costs; see [27, eq. (11)]. For $i = 1,\ldots,p$, we compute $D_u(q_i(\bar{u})^T\nabla_{\xi\xi}\widehat{J}(\bar{u},\bar{\mu})q_i(\bar{u}))$ using [27, eq. (20)], requiring the solution of the sensitivity and adjoint adjoint equations [27, eqs. (15), (18), (19)]. These PDE solves allow us to compute $D_u(\nabla\widehat{J}(\bar{u},\bar{\mu})^T s_\xi)$ using [27, eq. (14)] and $D_u(s_\xi^T\nabla_{\xi\xi}\widehat{J}(\bar{u},\bar{\mu})s_\xi)$ without additional costs.

To summarize, the evaluation of $\Psi(\bar{u};\rho)$ involves the solution of the state equation, and $2p+1$ solutions of linear equations, and that of $D_u\Psi(\bar{u};\rho)$ requires, in addition, $2p$ solutions of linear equations. In order to compute the Riesz representation of $D_u\Psi(\bar{u};\rho)$, we solve one more linear PDE.

7. Applications and Numerical Results. We formulate and analyze two distributionally robust optimization problems with PDEs, and present numerical results.

7.1. DRO of a Steady Burgers' Equation. We formulate an optimal control problem of a parameterized steady Burgers' equation that was studied in [33, 30, 28] for risk-averse objective functions other than (1.9). We consider

$$\min_{u\in U}\ \sup_{P\in\mathcal{P}}\ \mathbb{E}_P[(1/2)\|S(u,\xi)-y_d\|^2_{L^2(\mathsf{D})}] + (\alpha/2)\|u\|^2_{L^2(\mathsf{D})}, \tag{7.1}$$

where $\mathsf{D} = (0,1)$, $U = L^2(\mathsf{D})$, $\alpha = 10^{-3}$, $y_d = 1$, and $S(u,\xi) \in Y = H^1(\mathsf{D})$ solves the weak form of the steady Burgers' equation

$$\begin{aligned} -\nu(\xi)y_{xx}(x) + y(x)y_x(x) &= \xi_2/100 + u(x), \qquad x \in \mathsf{D},\\ y(0) = 1 + \xi_3/1000, &\quad y(1) = \xi_4/1000, \end{aligned} \tag{7.2}$$

where $p = 4$, $\xi \in \mathbb{R}^p$, $u \in U$ and $\nu : \mathbb{R}^p \to \mathbb{R}_{++}$, $\nu(\xi) = 10^{\xi_1 - 2}$. Deterministic control problems subject to (7.2) are analyzed in [49, 50].

We define $V = H_0^1(\mathsf{D})$ and $e = (e_1, e_2) : H^1(\mathsf{D}) \times H^{-1}(\mathsf{D}) \times \mathbb{R}^p \to V^* \times \mathbb{R}^2$ by

$$\langle e_1(y, u, \xi), v\rangle_{V^*,V} = \int_{\mathsf{D}} [\nu(\xi) y_x(x) v_x(x) + (y(x) y_x(x) - \tfrac{\xi_2}{100} - u(x)) v(x)] dx \tag{7.3}$$

for all $v \in V$ and $e_2(y, u, \xi) = (y(0) - 1 - \xi_3/1000, y(1) - \xi_4/1000)$; cf. [49, p. 79]. Using the derivations in [49, p. 81], we can show that e is twice continuously differentiable. The function $J : Y \times U \to \mathbb{R}_+$, $J(y, u) = (1/2)\|y - y_d\|^2_{L^2(\mathsf{D})} + (\alpha/2)\|u\|^2_{L^2(\mathsf{D})}$ is convex and infinitely times continuously differentiable. Hence, the differentiability requirements in Assumption 3.12 are met, and Assumption 3.2 holds.

The following considerations are based on [30, sect. 5.2.1] and [49, Chap. 4]. Let $\xi \in \mathbb{R}^p$ be arbitrary. Define $\varepsilon : \mathbb{R}^p \to \mathbb{R}$ by $\varepsilon(\xi) = \nu(\xi)/2$.[3] Using [49, Lem. 2.2], we can show that there exists $y_0(\xi) \in H^1(\mathsf{D})$ such that $e_2(y_0(\xi), u, \xi) = 0$ for all $u \in U$, and $\|y_0(\xi)\|_{L^2(\mathsf{D})} \le \varepsilon(\xi)$. The derivations in [49, sect. 2.3] imply that there exists a solution $S(u, \xi) \in Y$ of the weak form of (7.2). If $\nu(\xi)$ is sufficiently large, then the steady Burgers's equation (7.2) has a unique solution [49, Thm. 2.13].

Fix $u \in U_{\mathrm{ad}}$. Using [5, Thm. 8.2.9], we can show that the set-valued solution mapping $\mathcal{S}(u, \cdot) : \mathbb{R}^p \rightrightarrows Y$, $\mathcal{S}(u, \xi) = \{\, y \in Y : e(y, u, \xi) = 0 \,\}$ is measurable and there exists a measurable selection $S(u, \cdot)$ of $\mathcal{S}(u, \cdot)$ with $e(S(u, \xi), u, \xi) = 0$ for all $\xi \in \mathbb{R}^p$.

Let $(y, u, \xi) \in Y \times U \times \mathbb{R}^p$ fulfill $e(y, u, \xi) = 0$. The operator $e_y(y, u, \xi) \in \mathscr{L}(Y, V^* \times \mathbb{R}^2)$ is surjective [49, Thm. 3.3]. Combining the proof of [49, Thm. 3.3] and a Fredholm-type alternative, we can show that $e_y(y, u, \xi)$ is bijective. The bounded mapping theorem and the implicit function theorem imply that S is locally unique.

Let $S(u, \cdot)$ be a measurable selection of $\mathcal{S}(u, \cdot)$. We show that $\|S(u, \cdot)\|_Y^2$ is uniformly integrable. Using [49, Lem. 2.3] and the derivations in [49, p. 71], we obtain

$$\begin{aligned} \|S(u, \xi)\|_Y &\le (1 + \sqrt{8})\|y_0(\xi)\|_Y + (\sqrt{8}/\nu(\xi))\|u\|_{H^{-1}(\mathsf{D})} + (\sqrt{8}/\nu(\xi))\|y_0(\xi)\|_Y^2, \\ \|y_0(\xi)\|_Y^2 &\le (\varepsilon(\xi)^2((1 + 10^{-3}\xi_3)^2 + 10^{-6}\xi_4^2) + \varepsilon(\xi)^4)^2 + \varepsilon(\xi)^2. \end{aligned} \tag{7.4}$$

Combined with Lemmas 3.4, A.1, and A.3, we can show that the square of each addend in (7.4) is uniformly integrable. Jensen's inequality and the compact embedding $L^2(\mathsf{D}) \hookrightarrow H^{-1}(\mathsf{D})$, imply that $\|S(u, \cdot)\|_Y^2$ is uniformly integrable.

Discretization and Numerical Results. We transformed the Burgers' equation (7.2) to one with homogeneous boundary conditions, and discretized it using continuous piecewise linear finite elements on a uniform mesh of $(0, 1)$ with 2000 elements as in [30, sect. 5.2.2].

We approximated the DROP (7.1) with the DROP (1.5) and used Algorithm 2.1 to compute a stationary point of (1.5). We implemented Algorithm 2.1 in `Python` using `UFL` [4] to evaluate derivatives of J and e, and `FEniCs` [3] to compute the solutions to the PDEs (see section 6.2).

We chose the initial point $u_0 = 0$, and $(\tau_1, \nu_1, \eta_1) = 10^{-2}(1, 10^{-2}, 1)$ in Algorithm 2.1 and used $(\tau_{k+1}, \nu_{k+1}, \eta_{k+1}) = 10^{-1}(\tau_k, 10^{-1}\nu_k, \eta_k)$ to update the smoothing parameters. The parameter ν_k were decreased faster than τ_k and η_k due to the term $(2\nu_k)^{1/2}$ in (2.12). Algorithm 2.1 used `moola` [43] with its default settings except of using the termination tolerance 10^{-4} for each inner iteration of Algorithm 2.1, Wolfe line search, and `LBFGS`. The TRPs (2.8) were solved using the Moré–Sorensen algorithm [39].

[3] We choose $n = n(\xi) = ((1 + 10^{-3}\xi_3)^2 + 10^{-6}\xi_4^2)/(4\varepsilon(\xi)) + 1 \ge 1$ in the proof of [49, Lem. 2.2].

Table 1
Iteration history of Algorithm 2.1 applied to the approximated DROP of steady Burgers' equation (7.2), with $\Delta = \sigma_1 = 0.1$, $\sigma_0 = 0$, $\bar{\mu} = 0$, $\bar{\Sigma} = I$ and $t^k = (\tau_k, \nu_k, \eta_k)$.

k	$\widetilde{\Psi}(u^k; t^k)$	$\|\nabla_u \widetilde{\Psi}(u^k; t^k)\|_U$	#iter	$\frac{\|u^k - u^{k-1}\|_U}{1+\|u^{k-1}\|_U}$	$\#\widetilde{\Psi}(u^k; t^k)$	$\#\nabla_u \widetilde{\Psi}(u^k; t^k)$
1	7.97059e-03	6.13993e-05	18	8.24726e-01	21	21
2	4.71019e-03	9.30584e-05	9	7.27281e-02	11	11
3	4.54354e-03	8.85734e-05	3	3.23832e-03	5	5

Figure 1 depicts the controls u_N^* and u_{DR}^* and their corresponding states for three different ambiguity sets, where u_N^* is a stationary point of the nominal control problem (2.2) and u_{DR}^* is the final iterate of Algorithm 2.1. The robust controls depicted in Figure 1 have a similar structure to those obtained in [33, sect. 6.2] via the minimization of the conditional value-at-risk combined with a sample average approximation. For the approach in [33], each evaluation of the cost function and its gradient requires as many solutions of the Burgers' equation (7.2) as samples used, ranging from 19,000 to 23,000 in [33, sect. 6.2]. Our approach requires 37 solutions of (7.2) and 629 solutions of linear PDEs in total for the setup displayed in Table 1. We note that the cost function used here and in [33, sect. 6.2] are different and, hence, corresponding optimal controls cannot be compared directly. Our point is, however, that our approach produced reasonable controls with moderate computational costs.

We present detailed numerical results for the data $\Delta = \sigma_1 = 0.1$, $\sigma_0 = 0$, $\bar{\Sigma} = I$ and $\bar{\mu} = 0$. Table 1 provides an iteration history of Algorithm 2.1. It displays the objective function value of (2.1) and the U-norm of its gradient at the computed stationary point u^k of (2.1) for each outer iteration k of Algorithm 2.1. Moreover, it shows the number of inner iterations performed, a relative distance of subsequent stationary points, and the number of objective and gradient evaluations. The number of outer iterations of Algorithm 2.1 and the error of subsequent iterates of Algorithm 2.1 decrease monotonically.

Table 2 displays for the stationary points u_N^* and u_{DR}^* the statistics

$$
\begin{aligned}
\mathrm{E}^m(u) &= \max_{1\le i\le m} \mathbb{E}_{P_i}[\widehat{J}(u,\xi)], \quad \mathrm{SD}^m(u) = \max_{1\le i\le m} \mathrm{SD}_{P_i}[\widehat{J}(u,\xi)], \\
\mathrm{u}r\mathrm{SD}^m(u) &= \max_{1\le i\le m} \mathbb{E}_{P_i}\big[\big(\widehat{J}(u,\xi) - \mathbb{E}_{P_i}[\widehat{J}(u,\xi)]\big)_+^r\big], \quad r \in \{1,2\}, \\
Q_\beta^m(u) &= \max_{1\le i\le m} \mathrm{VaR}_{P_i,\beta}(\widehat{J}(u,\xi)),
\end{aligned}
\tag{7.5}
$$

where $P_i = \mathcal{N}(\hat{\mu}_i, \hat{\sigma}_i^2\bar{\Sigma}) \in \mathcal{P}$, $m \in \mathbb{N}$, SD_{P_i} is the standard deviation, $\mathrm{u}r\mathrm{SD}_{P_i}$ is the upper-r-semideviation and $\mathrm{VaR}_{P_i,\beta}$, $\beta \in (0,1)$, is the value-at-risk. Here, $\hat{\mu}_i$ are uniformly and independently distributed over $\{\, \mu \in \mathbb{R}^p : \|\mu\|_2 \le \Delta \,\}$ and $\hat{\sigma}_i$ on $[\sigma_0, \sigma_1]$ for $i = 1, \dots, m$. We chose $m = 10$ and approximated the quantities in (7.5) with 1000 independent samples. The statistics reported in Table 2 verify empirically that the distributionally robust control is more robust than the nominal control. We obtained similar results as in Table 2 for different choices of the parameters Δ, σ_0 and σ_1. The numerical results indicate that the objective function Ψ (see (1.6)) may be nonsmooth at $u_{\mathrm{DR}}^*(\Delta)$ for several different values of Δ.

7.2. DRO of an Unsteady Burgers' Equation. We consider the distributionally robust optimization of an unsteady Burgers' equation. We consider

$$
\min_{u\in U_{\mathrm{ad}}} \sup_{P\in\mathcal{P}} \mathbb{E}_P[(1/2)\|S(u,\xi) - y_d\|_X^2] + (\alpha/2)\|u_1\|_{L^2(\mathrm{I})}^2 + (\alpha/2)\|u_2\|_{L^2(\mathrm{I})}^2,
\tag{7.6}
$$

TABLE 2

Statistics (see (7.5)) for nominal control u_N^* and distributionally robust control $u_{\mathrm{DR}}^*(\Delta)$, associated to steady Burgers' equation (7.2), with $\Delta = \sigma_1 = 0.1$, $\sigma_0 = 0$, $\bar{\mu} = 0$, and $\bar{\Sigma} = I$.

u	$\mathrm{E}^m(u)$	$\mathrm{u1SD}^m(u)$	$\mathrm{SD}^m(u)$	$\mathrm{u2SD}^m(u)$	$Q^m_{.50}(u)$	$Q^m_{.80}(u)$	$Q^m_{.95}(u)$
u_N^*	5.27694e-03	3.83522e-02	3.36866e-03	2.81567e-03	3.90261e-03	8.68155e-03	1.12073e-02
u_{DR}^*	5.01929e-03	3.43171e-02	2.70053e-03	2.26590e-03	3.87501e-03	7.68191e-03	9.81026e-03

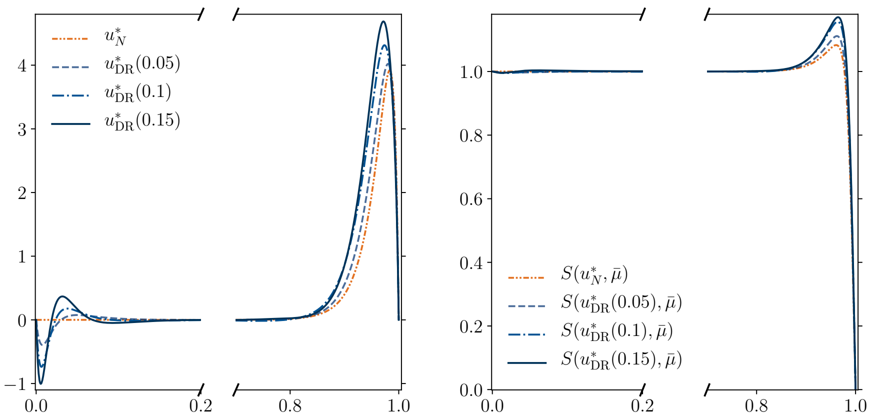

FIG. 1. *(Left) Stationary control u_N^* of (2.2) and $u_{\mathrm{DR}}^*(\Delta)$ of (1.5) for $\Delta = \sigma_1 = 0.05$, $\Delta = \sigma_1 = 0.1$ and $\Delta = \sigma_1 = 0.15$, associated to the approximated DROP of the steady Burgers' equation (7.2). The remaining data that defines the ambiguity set $\mathcal{P}$ (see (1.2)) is $\sigma_0 = 0$, $\bar{\mu} = 0$, $\bar{\Sigma} = I$. (Right) Corresponding states evaluated at $(u_N^*, \bar{\mu})$ and $(u_{\mathrm{DR}}^*(\Delta), \bar{\mu})$. Each graph is approximately constant between the breaks.*

where $U_{\mathrm{ad}} \subset U = L^2(\mathsf{I}) \times L^2(\mathsf{I})$, $\mathsf{I} = (0,1)$, $\mathsf{D} = (0,1)$, $\alpha = 0.01$, $y_d = 0.075$, $X = L^2(\mathsf{I}, L^2(\mathsf{D}))$, $Y = W(\mathsf{I}; L^2(\mathsf{D}), H^1(\mathsf{D}))$,[4] and $S(u,\xi)$ solves the weak form of the unsteady Burgers' equation

$$\begin{aligned} &y_t(x,t) = \nu(\xi) y_{xx}(x,t) - y(x,t)y_x(x,t) + (\xi_4/100)t, \quad x \in \mathsf{D}, \\ &y(x,0) = \phi(x,\xi), \quad x \in \mathsf{D}, \qquad y_x(0,t) = u_1(t), \quad y_x(1,t) = u_2(t), \quad t \in \mathsf{I}, \end{aligned} \tag{7.7}$$

where $(u_1, u_2) \in U$, $\nu : \xi \to \mathbb{R}_{++}$, $\nu(\xi) = 10^{\xi_1 - 1}$, and $\phi : \mathsf{D} \times \mathbb{R}^p \to \mathbb{R}$, $\phi(x,\xi) = (1 - 10\xi_2)x^2(1 + 10\xi_3 - x)(1-x)$. The model (7.7) is based on the one used in [11, sect. 7].

We show that Assumptions 3.1, 3.2, and 3.12.2 hold. We define $V = L^2(\mathsf{I}, H^1(\mathsf{D}))$, $e = (e_1, e_2) : Y \times U \times \mathbb{R}^p \to V^* \times L^2(\mathsf{D})$ and $e_2(y,u,\xi) = y(\cdot,0) - \phi(\cdot,\xi)$ by

$$\begin{aligned} \langle e_1(y,u), v\rangle_{V^*,V} = \langle y_t, v\rangle_{V^*,V} &+ \int_{\mathsf{I}} [u_1(t)v(0,t) - u_2(t)v(1,t)]dt \\ &+ \int_{\mathsf{I}}\int_{\mathsf{D}} [\nu(\xi)y_x(x,t)v_x(x,t) + (y(x,t)y_x(x,t) - 10^{-2}\xi_4 t)v(x,t)]dxdt, \end{aligned}$$

for all $v \in V$; cf. [49, p. 145].

Let $u \in U$ and $\xi \in \mathbb{R}^p$ be arbitrary. Based on the derivations in [49, p. 146], we can show that e is twice continuously differentiable. The function $J : Y \times U \to \mathbb{R}_+$, $J(y,u) = (1/2)\|y - y_d\|_X^2 + (\alpha/2)\|u_1\|^2_{L^2(\mathsf{I})} + (\alpha/2)\|u_2\|^2_{L^2(\mathsf{I})}$ is convex and infinitely times continuously differentiable. Hence, Assumption 3.2 holds. The state equation

[4]Here, $W(\mathsf{I}; L^2(\mathsf{D}), H^1(\mathsf{D})) = \{\, v \in L^2(\mathsf{I}, H^1(\mathsf{D})) : \, v_t \in L^2(\mathsf{I}, H^1(\mathsf{D})^*) \,\}$ [24, pp. 39–40].

$e(y,u,\xi) = 0$ has a unique solution $S(u,\xi)$ [51, Thm. 2.3] and $e_y(S(u,\xi),u,\xi) \in \mathscr{L}(Y,Z)$ is bijective [51, Prop. 2.5]. The bounded mapping theorem implies that $e_y(S(u,\xi),u,\xi)$ is boundedly invertible. Continuity of $S(u,\cdot)$ follows from the implicit function theorem. Hence, Assumption 3.12.2 holds. The proof of [51, Thm. 2.4] implies that $\{\,(y,u) \in Y \times U : e(y,u,\xi) = 0\,\}$ is weakly closed. From [51, Thm. 2.3] and its proof, we deduce that there exists $C(\xi) > 0$ such that

$$\|S(u,\xi)\|_Y \leq C(\xi)(1 + \|u_1\|_{L^2(\mathsf{I})} + \|u_2\|_{L^2(\mathsf{I})}), \quad \text{for all} \quad u = (u_1,u_2) \in U.$$

Hence, Lemma B.1 implies that $S(\cdot,\xi)$ is weakly-weakly continuous. Since $\xi \in \mathbb{R}^p$ is arbitrary, Assumption 3.1 holds.

We show that Assumption 3.3 is fulfilled. Let $(u,\xi) \in U \times \mathbb{R}^p$ be arbitrary. It suffices to prove that $\|S(u,\cdot)\|_V^2$ is uniformly integrable as $\|S(u,\xi)\|_X^2 \leq \|S(u,\xi)\|_V^2$. The proof of [49, Thm. 4.2] and [51, Prop. A.6] imply that

(7.8)
$$\|S(u,\xi)\|_V^2 \leq \frac{3|\xi_4|}{\nu(\xi)^2} + \frac{\|\phi(\cdot,\xi)\|_{L^2(\mathsf{D})}^2}{\nu(\xi)} + 6\|u_1\|_{L^2(\mathsf{I})}^2 + 6\|u_2\|_{L^2(\mathsf{I})}^2 + \frac{c_1 c_2(\xi)^2}{\nu(\xi)^4} + 2c_2(\xi),$$

where $c_1 > 0$ and $c_2 : \mathbb{R}^p \to \mathbb{R}_+$ is defined by $c_2(\xi) = 4(1 + \|\phi(\cdot,\xi)\|_{L^2(\mathsf{D})}^2)^2$. We have $\|\phi(\cdot,\xi)\|_{L^2(\mathsf{D})} = 630^{-1}(1-10\xi_2)^2(600\xi_3^2 + 45\xi_3 + 1)$. Using Young's inequality, and Lemmas 3.4, A.1, and A.3, we can show that each addend in (7.8) is uniformly integrable and, hence, $\|S(u,\cdot)\|_V^2$ is uniformly integrable. From Example 3.5, we deduce that Assumption 3.3 holds.

Discretization and Numerical Results. We discretized the unsteady Burgers' equation (7.7) in time, using the implicit Euler scheme on a uniform mesh of I with 100 time steps as in [49, sect. 5.6.1]. For the spatial discretization, we used piecewise linear finite elements on a uniform mesh of D with 100 elements. We used $U_{\text{ad}} = U$, approximated the DROP (7.6) by the DROP (1.5) and applied Algorithm 2.1 to (1.5). We chose $u^0 = u_N^*$, the stationary control of (2.2), and the same initial smoothing parameters and update rule as in section 7.1. In Algorithm 2.1, we used `scipy` with `LBFGS` with termination tolerance $< 10^{-2}$ for each inner iteration, and terminated Algorithm 2.1 when $\eta_k < 10^{-4}$.

Figure 2 depicts the stationary controls u_N^* of the nominal problem (2.2), and the distributionally robust controls u_{DR}^* of the approximated DROP of the unsteady Burgers' equation (7.7) for three ambiguity sets. Whereas the nominal control has a symmetric pattern, the robust controls are asymmetric, as a result of the non-symmetry of the parameterized initial condition ϕ (see (7.7)). The robust controls differ significantly from the nominal one. The statistics (see (7.5)) reported in Table 4 verify empirically that the distributionally robust control is more robust than the nominal one. We obtained similar numerical values for different choices of the data defining the ambiguity set. Table 3 provides an iteration history of Algorithm 2.1 applied to the approximated DROP of the unsteady Burgers' equation (7.7). The difference of successive iterates of the homotopy method converge to zero. The number of objective and gradient evaluations is quite low, and Algorithm 2.1 required only 53 solutions of the unsteady Burgers' equation (7.7) in total. Our numerical results indicate that the objective function Ψ (see (1.6)) may be nonsmooth at $u_{\text{DR}}^*(\Delta)$ for multiple values of Δ.

8. Conclusions. We developed a sampling-free approximation scheme for distributionally robust PDE-constrained optimization problems. Our approach incorporates second-order information from the reduced parameterized objective function (see

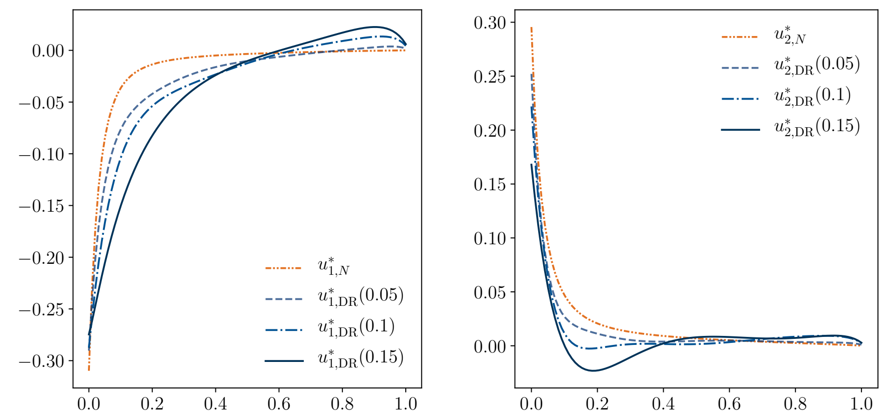

FIG. 2. *Stationary control* $(u^*_{1,N}, u^*_{2,N})$ *of* (2.2) *and* $(u^*_{1,\mathrm{DR}}(\Delta), u^*_{2,\mathrm{DR}}(\Delta))$ *of* (1.5) *for* $\Delta = 0.05 = 10\sigma_1$*,* $\Delta = 0.1 = 10\sigma_1$*, and* $\Delta = 0.5 = 10\sigma_1$*, associated to the approximated DROP of the unsteady Burgers' equation* (7.7)*. The remaining data that defines the ambiguity set* $\mathcal{P}$ *(see* (1.2)*) is* $\sigma_0 = 0$*,* $\bar{\mu} = 0$*,* $\bar{\Sigma} = I$*.*

TABLE 3
Iteration history of Algorithm 2.1 applied to the approximated DROP of the unsteady Burgers' equation (7.7)*, with for* $\Delta = 0.1$*,* $\sigma_0 = 0$*,* $\sigma_1 = 0.01$*,* $\bar{\Sigma} = I$ *and* $t^k = (\tau_k, \nu_k, \eta_k)$*.*

k	$\widetilde{\Psi}(u^k; t^k)$	$\\|\nabla_u \widetilde{\Psi}(u^k; t^k)\\|_U$	#iter	$\frac{\\|u^k - u^{k-1}\\|_U}{1+\\|u^{k-1}\\|_U}$	$\#\widetilde{\Psi}(u^k; t^k)$	$\#\nabla_u \widetilde{\Psi}(u^k; t^k)$
1	9.71222e-03	7.95245e-04	22	2.71162e-03	26	26
2	8.30158e-03	7.45890e-03	16	3.05599e-04	20	20
3	8.17309e-03	3.15171e-03	3	3.16757e-05	7	7

(1.3)) about the nominal parameter into the problem formulation (1.5), and only requires one solution of the state equation per evaluation of the surrogate cost function. We discussed why several other existing approaches for moment-based DROPs cannot be applied to the PDE-constrained DROPs studied here. We provided conditions for the PDE solution and cost function that ensure the existence of optimal solutions for the DROP (1.1), and the approximated and smoothed DROPs. To prove existence of worst-case distributions, we relied on the concept of uniform integrability. We applied our scheme to two nonlinear, nonconvex PDE-constrained problems and verified the assumptions required by our theory: the optimal control of a parameterized steady and unsteady Burgers' equations. The optimal control of the unsteady Burgers' equation is an important application in the field of PDE-constrained optimization since it allows the modeling of convection-diffusion phenomena [49, p. 69].

Future work includes the analysis of local convergence properties of the homotopy method, which may be built on the arguments used in [35], and the application of our scheme to further control problems with a larger number of parameters.

Appendix A. Bounds on Moments of Sub-Gaussian Random Vectors. We prove upper bounds on the strong moments of sub-Gaussian random vectors.

LEMMA A.1. *For all* $P \in \mathcal{P}$ *(see* (1.2)*) and each* $\gamma \geq 2$*, we have* $\mathbb{E}_P[\xi] \in \mathbb{R}^p$*, and*

$$\mathbb{E}_P[\\|\xi - \mathbb{E}_P[\xi]\\|_2^\gamma] \leq 2(\gamma/\mathrm{e})^{\gamma/2}(I \bullet \sigma_1\bar{\Sigma})^{\gamma/2}, \tag{A.1}$$

$$\mathbb{E}_P[\\|\xi\\|_2^\gamma] \leq 2^\gamma(\gamma/\mathrm{e})^{\gamma/2}(I \bullet \sigma_1\bar{\Sigma})^{\gamma/2} + 2^{\gamma-1}(\\|\bar{\Sigma}^{1/2}\\|_2\Delta + \\|\bar{\mu}\\|_2)^\gamma. \tag{A.2}$$

We apply Lemma A.2 to prove Lemma A.1.

Table 4

Statistics (see (7.5)) for nominal control u_N^* and distributionally robust control $u_{\mathrm{DR}}^*(\Delta)$, associated to the unsteady Burgers' equation (7.7), with $\Delta = 0.1$, $\sigma_0 = 0$, $\sigma_1 = 0.01$, $\bar{\mu} = 0$, and $\bar{\Sigma} = I$.

u	$\mathrm{E}^m(u)$	$\mathrm{u1SD}^m(u)$	$\mathrm{SD}^m(u)$	$\mathrm{u2SD}^m(u)$	$Q^m_{.50}(u)$	$Q^m_{.80}(u)$	$Q^m_{.95}(u)$
u_N^*	7.06471e-03	6.08569e-02	1.25907e-02	1.17093e-02	2.70676e-03	9.09762e-03	3.38810e-02
u_{DR}^*	6.56620e-03	5.78466e-02	1.14197e-02	1.06290e-02	3.05037e-03	8.45055e-03	3.06941e-02

Lemma A.2. *If $\xi : \Omega \to \mathbb{R}^p$ is a random vector with distribution $\mathbb{P} \in \mathcal{M}$ such that*

$$\mathbb{E}_{\mathbb{P}}[\xi] \in \mathbb{R}^p, \quad \text{and} \quad \ln \mathbb{E}_{\mathbb{P}}[\exp(d^T(\xi - \mathbb{E}_{\mathbb{P}}[\xi]))] \leq (1/2) d^T \Sigma d, \quad \text{for all} \quad d \in \mathbb{R}^p, \tag{A.3}$$

where $\Sigma \in \mathbb{S}^p_{++}$. Then, for all $\gamma \geq 2$,

$$\begin{aligned} \mathbb{E}_{\mathbb{P}}[\|\Sigma^{-1/2}(\xi - \mathbb{E}_{\mathbb{P}}[\xi])\|_2^\gamma] &\leq 2(\gamma/\mathrm{e})^{\gamma/2} p^{\gamma/2}, \\ \mathbb{E}_{\mathbb{P}}[\|\xi - \mathbb{E}_{\mathbb{P}}[\xi]\|_2^\gamma] &\leq 2(\gamma/\mathrm{e})^{\gamma/2} (I \bullet \sigma_1 \bar{\Sigma})^{\gamma/2}. \end{aligned} \tag{A.4}$$

Proof. Fix $\gamma \geq 2$, and define $\mathsf{W}, \mathsf{Z} : \Omega \to \mathbb{R}^p$ by $\mathsf{W}(\omega) = \xi(\omega) - \mathbb{E}_{\mathbb{P}}[\xi]$ and $\mathsf{Z}(\omega) = \Sigma^{-1/2}\mathsf{W}(\omega)$, respectively. We have $\mathbb{E}_{\mathbb{P}}[\mathsf{W}] = \mathbb{E}_{\mathbb{P}}[\mathsf{Z}] = 0$. Minkowski's inequality yields

$$\mathbb{E}_{\mathbb{P}}[\|\mathsf{Z}\|_2^\gamma] \leq \Big(\sum_{i=1}^p (\mathbb{E}_{\mathbb{P}}[|\mathsf{Z}_i|^\gamma])^{2/\gamma} \Big)^{\gamma/2}, \quad \mathbb{E}_{\mathbb{P}}[\|\mathsf{W}\|_2^\gamma] \leq \Big(\sum_{i=1}^p (\mathbb{E}_{\mathbb{P}}[|\mathsf{W}_i|^\gamma])^{2/\gamma} \Big)^{\gamma/2}. \tag{A.5}$$

Using (A.3), we obtain $\ln \mathbb{E}_{\mathbb{P}}[\exp(d^T \mathsf{Z})] \leq (1/2) d^T d$ for all $d \in \mathbb{R}^p$. Hence, W and Z are sub-Gaussian [10, Def. 7.1.1 (p. 185)]. For $i = 1, \ldots, p$, we obtain from [10, Lem. 7.1.4 (p. 187) and Lem. 1.1.4 (p. 7)] that

$$\mathbb{E}_{\mathbb{P}}[|\mathsf{Z}_i|^\gamma] \leq 2(\gamma/\mathrm{e})^{\gamma/2}, \quad \text{and} \quad \mathbb{E}_{\mathbb{P}}[|\mathsf{W}_i|^\gamma] \leq 2(\gamma/\mathrm{e})^{\gamma/2} (\sigma_1 \bar{\Sigma})_{ii}^{\gamma/2}.$$

Combined with (A.5), we deduce (A.4). ∎

Proof of Lemma A.1. Fix $P \in \mathcal{P}$. The definition of $\mathcal{P}$ (see (1.2)) implies that $\|\bar{\Sigma}^{-1/2}(\mathbb{E}_P[\xi] - \bar{\mu})\|_2 \leq \Delta$. We have $\mathbb{E}_P[\xi] \in \mathbb{R}^p$ because

$$\|\mathbb{E}_P[\xi]\|_2 \leq \|\bar{\Sigma}^{-1/2}\|_2 \|\bar{\Sigma}^{-1/2}(\mathbb{E}_P[\xi] - \bar{\mu})\|_2 + \|\bar{\mu}\|_2 \leq \|\bar{\Sigma}^{-1/2}\|_2 \Delta + \|\bar{\mu}\|_2 < \infty. \tag{A.6}$$

Now, the estimate (A.1) follows from Lemma A.2 (see (A.4)).

Next, we establish (A.2). The monotonicity and convexity of $\mathbb{R}_+ \ni z \mapsto z^\gamma$ yield

$$\|\xi\|_2^\gamma \leq (\|\xi - \mathbb{E}_P[\xi]\|_2 + \|\mathbb{E}_P[\xi]\|_2)^\gamma \leq 2^{\gamma-1}(\|\xi - \mathbb{E}_P[\xi]\|_2^\gamma + \|\mathbb{E}_P[\xi]\|_2^\gamma) \tag{A.7}$$

for every $\xi \in \mathbb{R}^p$. Now, (A.1), (A.6) and (A.7) yield (A.2). ∎

The next lemma is used to verify Assumption 3.3 in section 7.

Lemma A.3. *Define $a : \mathbb{R}^p \to \mathbb{R}$ by $a(\xi) = \sum_{i=1}^p \alpha_i \xi_i^{r_i}$, and $b : \mathbb{R}^p \to \mathbb{R}$ by $b(\xi) = \exp\big(\sum_{i=1}^p \beta_i \xi_i\big)$, where α_i, $\beta_i \in \mathbb{R}$ and $r_i \geq 1$ are fixed. Then, $|a|^r$, $|b|^s$ and $|a^t b^s|^r$ are uniformly integrable for all r, $t > 0$ and $s \in \mathbb{R}$.*

Proof. Fix $\xi \in \mathbb{R}^p$, r, $t > 0$, and $s \in \mathbb{R}$. We define $\bar{r} = \max\{1, r\} \geq 1$. The function $\mathbb{R}_+ \ni z \mapsto z^{\bar{r}}$ is convex and, hence, Jensen's and Young's inequality yield

$$\begin{aligned} |a(\xi)|^{r(2\bar{r}/r)} = |a(\xi)|^{2\bar{r}} &\leq p^{2\bar{r}-1} \sum_{i=1}^p |\alpha_i|^{2\bar{r}} |\xi_i|^{2\bar{r} r_i} \leq p^{2\bar{r}-1} \sum_{i=1}^p |\alpha_i|^{2\bar{r}} \|\xi\|_2^{2\bar{r} r_i}, \\ |a(\xi)^t b(\xi)^s|^{2\bar{r}} &\leq |a(\xi)|^{2t\bar{r}} |b(\xi)|^{2s\bar{r}} \leq (1/2)|a(\xi)|^{4t\bar{r}} + (1/2)|b(\xi)|^{4s\bar{r}}. \end{aligned} \tag{A.8}$$

Now, Lemmas 3.4 and A.1, and (A.8) imply the uniform integrability of $|a(\cdot)|^r$ for all $r > 0$. We have $|b(\xi)|^{2s} = \exp\left(\sum_{i=1}^{p} 2s\alpha_i\xi_i\right)$. The proof of Lemma 3.10 shows that $|b(\cdot)|^s$ is uniformly integrable for all $s \in \mathbb{R}$. The first two assertions, Lemma 3.4 and (A.8) imply uniform integrability of $|a(\cdot)^t b(\cdot)^s|^r$ for each r, $t > 0$ and $s \in \mathbb{R}$. ∎

Appendix B. Weak-Weak Continuity of Solution Operators. We provide conditions implying weak-weak continuity of the solution operator of a PDE.

LEMMA B.1. *Let Y, U and Z be Banach spaces, Y be reflexive, and $U_{ad} \subset U$. For each $u \in U_{ad}$, let $S(u) \in Y$ be the unique solution to: Find $y \in Y$: $e(y, u) = 0$, where $S : U_{ad} \to Y$ and $e : Y \times U \to Z$. Suppose that $\rho : \mathbb{R} \to \mathbb{R}$ is monotonically increasing with $\|S(u)\|_Y \le \rho(\|u\|_U)$ for all $u \in U_{ad}$, and that $X = \{\, (y, u) \in Y \times U_{ad} : e(y, u) = 0 \,\}$ is weakly sequentially closed. Then, S is weakly-weakly continuous.*

We apply Lemma B.2, which is essentially known, to prove Lemma B.1.

LEMMA B.2. *Let X be a Banach space, $x \in X$, and $(x_k) \subset X$. If every subsequence of (x_k) has a subsequence that converges weakly to x, then $x_k \rightharpoonup x$ as $k \to \infty$.*

Proof. Fix $f \in X^*$. By assumption, every subsequence of $(\langle f, x_k\rangle_{X^*,X}) \subset \mathbb{R}$ has a subsequence that converges to $\langle f, x\rangle_{X^*,X} \in \mathbb{R}$. Hence $\langle f, x_k\rangle_{X^*,X} \to \langle f, x\rangle_{X^*,X}$. ∎

Proof of Lemma B.1. Fix $u \in U_{\mathrm{ad}}$ and $(u_k) \subset U_{\mathrm{ad}}$ with $u_k \rightharpoonup u$ as $k \to \infty$. The boundedness of (u_k) implies that of $(S(u_k))$. Let $(S(u_k))_K$ be any subsequence of $(S(u_k))$. By boundedness and reflexivity, there exist $y \in Y$ and a subsequence $(S(u_k))_{K'}$ of $(S(u_k))_K$ such that $S(u_k) \rightharpoonup y$ as $K' \ni k \to \infty$. Since $(S(u_k), u_k) \subset X$ and X is weakly sequentially closed, we get $(y, u) \in X$ and, hence, $y = S(u)$. Thus, every subsequence of $(S(u_k))$ has a further subsequence converging to $S(u)$ weakly. Lemma B.2 yields $S(u_k) \rightharpoonup S(u)$ as $k \to \infty$. ∎

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*, SIAM/ASA J. Uncertainty Quantification, 5 (2017), pp. 1166–1192, https://doi.org/10.1137/16M106306X.

[2] A. Alla, M. Hinze, P. Kolvenbach, O. Lass, and S. Ulbrich, *A certified model reduction approach for robust parameter optimization with PDE constraints*, Adv. Comput. Math., 45 (2019), pp. 1221–1250, https://doi.org/10.1007/s10444-018-9653-1.

[3] M. S. Alnæs, J. Blechta, J. Hake, A. Johansson, B. Kehlet, A. Logg, C. Richardson, J. Ring, M. E. Rognes, and G. N. Wells, *The FEniCS Project Version 1.5*, Archive of Numerical Software, 3 (2015), pp. 9–23, https://doi.org/10.11588/ans.2015.100.20553.

[4] M. S. Alnæs, A. Logg, K. B. Ølgaard, M. E. Rognes, and G. N. Wells, *Unified Form Language: A Domain-specific Language for Weak Formulations of Partial Differential Equations*, ACM Trans. Math. Softw., 40 (2014), pp. 9:1–9:37, https://doi.org/10.1145/2566630.

[5] J.-P. Aubin and H. Frankowska, *Set-Valued Analysis*, Modern Birkhäuser Classics, Springer, Boston, MA, 2009, https://doi.org/10.1007/978-0-8176-4848-0.

[6] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski, *Robust Optimization*, Princeton Series in Applied Mathematics, Princeton University Press, Princeton, NJ, 2009.

[7] P. Billingsley, *Probability and Measure*, Wiley Series in Probability and Statistics, John Wiley & Sons, Hoboken, NJ, 2012. Anniversary Edition.

[8] V. I. Bogachev, *Weak Convergence of Measures*, vol. I & II, American Mathematical Society, Providence, RI, 2018.

[9] J. F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer Series in Operations Research, Springer, New York, NY, 2000, https://doi.org/10.1007/978-1-4612-1394-9.

[10] V. V. Buldygin and Yu. V. Kozachenko, *Metric Characterization of Random Variables and Random Processes*, vol. 188 of Translations of Mathematical Monographs, American Mathematical Society, Providence, RI, 2000.

[11] C. Büskens and R. Griesse, *Parametric sensitivity analysis of perturbed PDE optimal control problems with state and control constraints*, J. Optim. Theory Appl., 131 (2006), pp. 17–35, https://doi.org/10.1007/s10957-006-9122-8.

[12] P. Chen, U. Villa, and O. Ghattas, *Taylor approximation for PDE-constrained optimization under uncertainty: Application to turbulent jet flow*, Proc. Appl. Math. Mech., 18 (2018), p. e201800466, https://doi.org/10.1002/pamm.201800466.

[13] X. Chen, Z. Nashed, and L. Qi, *Smoothing methods and semismooth methods for nondifferentiable operator equations*, SIAM J. Numer. Anal., 38 (2000), pp. 1200–1216, https://doi.org/10.1137/S0036142999356719.

[14] Z. Chen, M. Sim, and H. Xu, *Distributionally Robust Optimization with Infinitely Constrained Ambiguity Sets*, Oper. Res., (2019), https://doi.org/10.1287/opre.2018.1799.

[15] S. Conti, H. Held, M. Pach, M. Rumpf, and R. Schultz, *Shape Optimization Under Uncertainty—A Stochastic Programming Perspective*, SIAM J. Optim., 19 (2009), pp. 1610–1632, https://doi.org/10.1137/070702059.

[16] S. Conti, H. Held, M. Pach, M. Rumpf, and R. Schultz, *Risk Averse Shape Optimization*, SIAM J. Control Optim., 49 (2011), pp. 927–947, https://doi.org/10.1137/090754315.

[17] F. de Gournay, G. Allaire, and F. Jouve, *Shape and topology optimization of the robust compliance via the level set method*, ESAIM Control. Optim. Calc. Var., 14 (2008), pp. 43–70, https://doi.org/10.1051/cocv:2007048.

[18] E. Delage and Y. Ye, *Distributionally Robust Optimization Under Moment Uncertainty with Application to Data-Driven Problems*, Oper. Res., 58 (2010), pp. 595–612, https://doi.org/10.1287/opre.1090.0741.

[19] R. Gao and A. J. Kleywegt, *Distributionally robust stochastic optimization with Wasserstein distance*, (2016), https://arxiv.org/abs/1604.02199.

[20] S. Garreis and M. Ulbrich, *Constrained optimization with low-rank tensors and applications to parametric problems with PDEs*, SIAM J. Sci. Comput., 39 (2017), pp. A25–A54, https://doi.org/10.1137/16M1057607.

[21] L. Hertlein and M. Ulbrich, *An Inexact Bundle Algorithm for Nonconvex Nonsmooth Minimization in Hilbert Space*, SIAM J. Control Optim., 57 (2019), pp. 3137–3165, https://doi.org/10.1137/18M1221849.

[22] R. Herzog and F. Schmidt, *Weak lower semi-continuity of the optimal value function and applications to worst-case robust optimal control problems*, Optimization, 61 (2012), pp. 685–697, https://doi.org/10.1080/02331934.2011.603322.

[23] E. Hille and R. S. Phillips, *Functional Analysis and Semi-Groups*, vol. 31 of Colloquium Publications, American Mathematical Society, Providence, RI, 1957.

[24] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich, *Optimization with PDE Constraints*, vol. 23 of Mathematical Modelling: Theory and Applications, Springer, Dordrecht, 2009, https://doi.org/10.1007/978-1-4020-8839-1.

[25] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 2nd ed., 2013, https://doi.org/10.1017/9781139020411.

[26] O. Kallenberg, *Foundations of Modern Probability*, Probability and its Applications, Springer, New York, NY, 1997, https://doi.org/10.1007/b98838.

[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, https://doi.org/10.1007/s11081-018-9388-3.

[28] D. P. Kouri, *A Multilevel Stochastic Collocation Algorithm for Optimization of PDEs with Uncertain Coefficients*, SIAM/ASA J. Uncertainty Quantification, 2 (2014), pp. 55–81, https://doi.org/10.1137/130915960.

[29] D. P. Kouri, *A Measure Approximation for Distributionally Robust PDE-Constrained Optimization Problems*, SIAM J. Numer. Anal., 55 (2017), pp. 3147–3172, https://doi.org/10.1137/15M1036944.

[30] D. P. Kouri, M. Heinkenschloss, D. Ridzal, and B. van Bloemen Waanders, *A Trust-Region Algorithm with Adaptive Stochastic Collocation for PDE Optimization under Uncertainty*, SIAM J. Sci. Comput., 35 (2013), pp. A1847–A1879, https://doi.org/10.1137/120892362.

[31] D. P. Kouri, M. Heinkenschloss, D. Ridzal, and B. G. van Bloemen Waanders, *Inexact objective function evaluations in a trust-region algorithm for PDE-constrained optimization under uncertainty*, SIAM J. Sci. Comput., 36 (2014), pp. A3011–A3029,

https://doi.org/10.1137/140955665.
[32] D. P. Kouri and A. Shapiro, *Optimization of PDEs with Uncertain Inputs*, in Frontiers in PDE-Constrained Optimization, H. Antil, D. P. Kouri, M.-D. Lacasse, and D. Ridzal, eds., vol. 163 of The IMA Volumes in Mathematics and its Applications, Springer, New York, NY, 2018, pp. 41–81, https://doi.org/10.1007/978-1-4939-8636-1_2.
[33] D. P. Kouri and T. Surowiec, *Risk-Averse PDE-Constrained Optimization Using the Conditional Value-At-Risk*, SIAM J. Optim., 26 (2016), pp. 365–396, https://doi.org/10.1137/140954556.
[34] D. P. Kouri and T. Surowiec, *Existence and Optimality Conditions for Risk-Averse PDE-Constrained Optimization*, SIAM/ASA J. Uncertainty Quantification, 6 (2018), pp. 787–815, https://doi.org/10.1137/16M1086613.
[35] D. P. Kouri and T. M. Surowiec, *Epi-Regularization of Risk Measures*, Math. Oper. Res., (2019), https://doi.org/10.1287/moor.2019.1013.
[36] 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, https://doi.org/10.1137/16M108269X.
[37] A. S. Lewis and H. S. Sendov, *Twice differentiable spectral functions*, SIAM J. Matrix Anal. Appl., 23 (2001), pp. 368–386, https://doi.org/10.1137/S089547980036838X.
[38] J. Milz and M. Ulbrich, *An approximation scheme for distributionally robust nonlinear optimization.* accepted for publication in SIAM J. Optim, 2020, https://www-m1.ma.tum.de/foswiki/pub/M1/Lehrstuhl/PublikationenUlbrich/MilzUlbrich-DRO.pdf.
[39] J. J. Moré and D. C. Sorensen, *Computing a Trust Region Step*, SIAM J. Sci. and Stat. Comput., 4 (1983), pp. 553–572, https://doi.org/10.1137/0904038.
[40] Yu. Nesterov, *Smoothing technique and its applications in semidefinite optimization*, Math. Program., 110 (2007), pp. 245–259, https://doi.org/10.1007/s10107-006-0001-8.
[41] G. Ch. Pflug and D. Wozabal, *Ambiguity in portfolio selection*, Quant. Financ., 7 (2007), pp. 435–442, https://doi.org/10.1080/14697680701455410.
[42] A. Ruszczyński and A. Shapiro, *Optimization of Risk Measures*, in Probabilistic and Randomized Methods for Design under Uncertainty, G. Calafiore and F. Dabbene, eds., Springer, London, 2006, pp. 119–157, https://doi.org/10.1007/1-84628-095-8_4.
[43] T. Schwedes, D. A. Ham, S. W. Funke, and M. D. Piggott, *Mesh Dependence in PDE-Constrained Optimisation*, SpringerBriefs in Mathematics of Planet Earth, Springer, Cham, 2017, https://doi.org/10.1007/978-3-319-59483-5.
[44] A. Shapiro, *Distributionally robust stochastic programming*, SIAM J. Optim., 27 (2017), pp. 2258–2275, https://doi.org/10.1137/16M1058297.
[45] A. Shapiro, D. Dentcheva, and A. Ruszczyńsk, *Lectures on Stochastic Programming: Modeling and Theory*, vol. 9 of MPS-SIAM Series on Optimization, SIAM, Philadelphia, PA, 2009, https://doi.org/10.1137/1.9780898718751.
[46] A. Shapiro and A. Kleywegt, *Minimax analysis of stochastic problems*, Optim. Methods Softw., 17 (2002), pp. 523–542, https://doi.org/10.1080/1055678021000034008.
[47] A. M.-C. So, *Moment inequalities for sums of random matrices and their applications in optimization*, Math. Program., 130 (2011), pp. 125–151, https://doi.org/10.1007/s10107-009-0330-5.
[48] A. Van Barel and S. Vandewalle, *Robust Optimization of PDEs with Random Coefficients Using a Multilevel Monte Carlo Method*, SIAM/ASA J. Uncertainty Quantification, 7 (2019), pp. 174–202, https://doi.org/10.1137/17M1155892.
[49] S. Volkwein, *Mesh-Independence of an Augmented Lagrangian-SQP Method in Hilbert Spaces and Control Problems for the Burgers Equation*, Dissertation, Technical University of Berlin, Berlin, 1997, https://imsc.uni-graz.at/volkwein/diss.ps.
[50] S. Volkwein, *Application of the Augmented Lagrangian-SQP Method to Optimal Control Problems for the Stationary Burgers Equation*, Comput. Optim. Appl., 16 (2000), pp. 57–81, https://doi.org/10.1023/A:1008777520259.
[51] S. Volkwein, *Second-order conditions for boundary control problems of the Burgers equation*, Control Cybernet., 30 (2001), pp. 249–278.
[52] W. Wiesemann, D. Kuhn, and M. Sim, *Distributionally Robust Convex Optimization*, Oper. Res., 62 (2014), pp. 1358–1376, https://doi.org/10.1287/opre.2014.1314.
[53] Y. Xu, W. Sun, and L. Qi, *A feasible direction method for the semidefinite program with box constraints*, Appl. Math. Lett., 24 (2011), pp. 1874–1881, https://doi.org/10.1016/j.aml.2011.05.010.