PERFORMANCE BOUNDS FOR PDE-CONSTRAINED OPTIMIZATION UNDER UNCERTAINTY[*]

PENG CHEN[†] AND JOHANNES O. ROYSET[‡]

Abstract. Computational approaches to PDE-constrained optimization under uncertainty may involve finite-dimensional approximations of control and state spaces, sample average approximations of measures of risk and reliability, smooth approximations of nonsmooth functions, penalty approximations of constraints, and many other kinds of inaccuracies. In this paper, we analyze the performance of controls obtained by an approximation-based algorithm and in the process develop estimates of optimality gaps for general optimization problems defined on metric spaces. Under mild assumptions, we establish that limiting controls have arbitrarily small optimality gaps provided that the inaccuracies in the various approximations vanish. We carry out the analysis for a broad class of problems with multiple expectation, risk, and reliability functions involving PDE solutions and appearing in objective as well as constraint expressions. In particular, we address problems with buffered failure probability constraints approximated via an augmented Lagrangian. We demonstrate the framework on an elliptic PDE with a random coefficient field and a distributed control function.

Key words. PDE-constrained optimization, stochastic optimization, uncertainty quantification, performance bounds, sample average approximation, smooth approximation, buffered probability

MSC codes. 65C20, 65D32, 65N12, 49J20, 93E20

DOI. 10.1137/21M1457916

1. Introduction. Optimization problems involving partial differential equations (PDEs) arise widely in design and control of physical systems [17, 41]. Recent applications include shape optimization [18, 14], turbulent combustion [11], acoustic wave propagation [43], metamaterials [9], and plasma fusion [42]. Problems of this kind usually come with significant uncertainty in the form of unknown external loadings, material coefficients, boundary or initial conditions, geometries, and other factors, which leads to PDE-constrained optimization problems under uncertainty. Formulations of these problems leverage scalarizations of the random quantities of interest via expectations, variances [1], superquantiles (a.k.a. conditional/average value-at-risk) [20, 13], worst-case measures of risk [24], and probabilities of failure [8]. These formulations tend to involve integration with respect to a probability measure, which needs to be approximated by numerical integration, such as Monte Carlo and quasi Monte Carlo techniques [35, 26, 15], sparse grid quadrature methods [10], or variance-reduction techniques based on Taylor expansions [11]. These approximations compound the already significant challenges associated with discretization and solution of the underlying PDE. Techniques to mitigate the overall computational cost include

[†]Oden Institute for Computational Engineering and Sciences, The University of Texas at Austin, Austin, TX 78712 USA (peng@oden.utexas.edu).

[‡]Operations Research Department, Naval Postgraduate School, Monterey, CA 93943 USA (joroyset@nps.edu).

those based on model reduction [6, 24], multilevel [7, 2] and multifidelity [27] approximations, low-rank tensor decomposition [3], stochastic collocation [19], stochastic Galerkin [23], and Taylor approximation [1]. In this paper, we analyze a multitude of approximations in PDE-constrained optimization under uncertainty. We focus on a broad class of problems involving several quantities of interest and, thus, several expectation, risk, and reliability functions. These quantities may arise in an objective function, as constraint functions, or both.

In addition to discretization of state and control spaces and numerical evaluation of integrals, the broad class of problems gives rise to several other approximations. Smooth functions approximate nonsmooth ones to facilitate the use of gradient-type algorithms. Penalty and Lagrangian terms model constraints. Weights aggregate multiple objective functions. The combined effect of this multitude of approximations on the obtained solutions is nontrivial and increasingly complex as we move towards more sophisticated optimization problems involving novel performance criteria and numerous quantities of interest. We lay out conditions under which small errors in the various approximations indeed lead to a guaranteed good performance of a control computed using these approximations. This does not hold in general. It is well known that approximating functions may converge pointwise to an actual function but have minimizers and minima far from those of the actual problem [38, Figure 4.5]. More concretely, a convergent Runga–Kutta method employed to solve an ordinary differential equation might still induce errors that prevent minimizers of the approximating optimization problems from converging to a minimizer of the actual problem [40, 16]. We present a unified framework that addresses a wide array of approximations and applications. It stretches beyond PDE-constrained optimization, but also specializes to an example with a buffered failure probability constraint and to problems defined in terms of an elliptic PDE with a random coefficient field and a distributed control function.

The various inaccuracies lead to approximating optimization problems that subsequently need to be solved by optimization algorithms. Since these problems are rarely convex, one cannot expect their minimizers to be within reach. We furnish performance bounds and optimality gaps for *any* solution—local, global, or neither—obtained from the approximating problems. Still, better solutions of the approximating problems naturally translate into tighter bounds.

The reasoning towards performance bounds and optimality gaps follows a novel breakdown of errors into those caused by discretization of the control space on one side and those produced by all other approximations on the other side. This division is motivated by the fact that the former imposes a restriction on the problem while the latter can swing either way. As a result, we often analyze the "other approximations" in a setting of finite dimensions, and this reduces the required assumptions. We omit an analysis of optimality conditions and the assumptions required to ensure convergence of stationary points of approximating problems to those of the actual problem; see, for example, [29, 21] for efforts in this direction. Our results provide the foundation for numerous algorithms, including those involving adaptive refinements. However, we defer detailed algorithms as well as numerical demonstrations to future publications.

Our main technical tool is epi-convergence; see, for example, [36, Chapter 7], [38, section 4.C], and references therein. It is well known that epi-convergence of approximating functions ensures that the corresponding minimizers can only converge to a minimizer of the limiting function. However, epi-convergence can be difficult to verify or might simply not hold for realistic problems involving a multitude of

approximations. Many studies concentrate on a single source of approximation; [22] deals with nonsmoothness and [12, 4] with sampling. The challenge of combining discretization of an underlying space with sample averages is illustrated by the less-than-ideal convergence results of Theorem 3.14 in [37], which anyhow fails to address other approximations and expectation constraints. The difficulties compound for PDE-constrained optimization problems because the approximating problems solved by numerical algorithms are defined on finite-dimensional Euclidean spaces while the actual problem resides in infinite dimensions. The main novelty of the present paper stems from our ability to handle a multitude of approximations. The challenge of establishing epi-convergence in such settings is well known. Even a sum rule does not hold in general: For extended real-valued functions f^ν, f, g^ν, g defined on a metric space, f^ν epi-converging to f and g^ν epi-converging to g do not ensure that $f^\nu + g^\nu$ epi-converges to $f + g$; see, e.g., the discussion in [38, section 4.C].

The rest of the paper is organized as follows: Section 2 considers formulations and illustrative examples. Section 3 develops a general approximation theory and, thus, lays the foundation for addressing PDE-constrained optimization problems in section 4. Section 5 demonstrates the theory in the context of optimal control of an elliptic PDE with a random coefficient field and two quantities of interest.

2. Problem formulations. We consider the stochastic optimization problem

$$(2.1) \qquad \underset{z \in Z}{\text{minimize}}\ \varphi(z) = \iota_A(z) + f(z) + h\Big(\mathbb{E}\big[G(\boldsymbol{\xi}, z)\big]\Big),$$

referred to as the *actual problem*, where the *control* z resides in a separable Banach space $(Z, \|\cdot\|_Z)$ and is restricted to an *admissible set* $A \subset Z$ as expressed by the *indicator function*; for any set C, $\iota_C(c) = 0$ if $c \in C$ and $\iota_C(c) = \infty$ otherwise. The *objective function* $\varphi: Z \to \overline{\mathbb{R}} = [-\infty, \infty]$ is further given by a *cost function* $f: Z \to \mathbb{R}$. The main complications stem from the last term in the objective function, which is defined by a *random field* $\boldsymbol{\xi}$, an m-dimensional vector of *quantities of interest* $G(\boldsymbol{\xi}, z)$, and a *monitoring function* $h: \mathbb{R}^m \to \overline{\mathbb{R}}$. The monitoring function assesses the various quantities of interest, which in turn depend implicitly on the solution of a *PDE* defined on a domain $D \subset \mathbb{R}^d$ and parametrized by $z \in Z$ and realizations of $\boldsymbol{\xi}$.

Formally, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and Ξ be a Hilbert space of functions on D equipped with the Borel σ-field $\mathcal{B}$. On the probability space, we define a random field, i.e., a measurable function $\boldsymbol{\xi}: \Omega \to \Xi$. We denote by ξ a realization of $\boldsymbol{\xi}$, i.e., $\xi = \boldsymbol{\xi}(\omega) \in \Xi$ for some $\omega \in \Omega$. Thus, boldface letters indicate a random element and regular font its realization. In the usual way, $\boldsymbol{\xi}$ defines another probability space $(\Xi, \mathcal{B}, P)$, where $P(B) = \mathbb{P}(\{\omega \in \Omega \mid \boldsymbol{\xi}(\omega) \in B\})$ for $B \in \mathcal{B}$. Without imposing any practical limitation, we let this probability space be complete.

We assume that the solutions of the PDE lie in a separable Banach space $(U, \|\cdot\|_U)$ and, for given $\xi \in \Xi$ and $z \in Z$, the PDE has a unique solution in U denoted by $s(\xi, z)$. Section 5 furnishes details about the existence of such solutions. For the majority of the development, it suffices to recognize that the solutions of the PDE for various ξ and z are given by a *solution mapping* $s: \Xi \times Z \to U$.

The quantities of interest are defined by the *performance functions* $g_i: U \times Z \to \mathbb{R}$, $i = 1, \ldots, m$, and the solution mapping s:

$$(2.2) \qquad G: \Xi \times Z \to \mathbb{R}^m, \quad \text{with } G(\xi, z) = \Big(g_1\big(s(\xi, z), z\big), \ldots, g_m\big(s(\xi, z), z\big)\Big).$$

For any $z \in Z$, we write $\mathbb{E}[G(\boldsymbol{\xi}, z)] = \int G(\xi, z) dP(\xi)$, which is assumed to be well defined and finite as discussed below.

We adopt the usual rules for extended arithmetic. In particular, $\infty - \infty = \infty$; see [38, section 1.D]. Thus, $z \notin A$ and/or $h(\mathbb{E}[G(\boldsymbol{\xi}, z)]) = \infty$ produce $\varphi(z) = \infty$, which im-

plies that z is infeasible. With the convention $\operatorname{argmin} \varphi = \{z \in Z \mid \varphi(z) = \inf \varphi < \infty\}$, where $\inf \varphi = \inf\{\varphi(z) \mid z \in Z\}$, this set of minimizers cannot include such z.

2.1. Illustrative examples. The actual problem (2.1) addresses a vast array of applications. We highlight some possibilities.

Example 2.1 (expectation minimization with regularization). Suppose that we seek to determine a control $z \in A \subset Z = L^2(D; \mathbb{R}^q)$, the space of Lebesgue square-integrable functions from D to $\mathbb{R}^q$, that minimizes the expected system performance as quantified by $g_1 : U \times Z \to \mathbb{R}$. This leads to the problem

$$\underset{z \in A}{\text{minimize}} \ \mathbb{E}\Big[g_1\big(s(\boldsymbol{\xi}, z), z\big)\Big] + \theta\|z\|_Z^2,$$

where a regularization term is included with $\theta \in [0, \infty)$. The problem is of the form (2.1): set $m = 1$, $h(w) = w$, *and* $f(z) = \theta\|z\|_Z^2$.

Example 2.2 (tracking objective under expectation constraint). While using the same admissible controls as in Example 2.1, suppose that we seek to bring the random state $s(\boldsymbol{\xi}, z)$ as close as possible to a target state $\bar{u} \in U$ on average and force a performance function $g_1 : U \times L^2(D; \mathbb{R}^q) \to \mathbb{R}$ to lie between α and β on average. This leads to the problem

$$\underset{z \in A}{\text{minimize}} \ \mathbb{E}\Big[\big\|s(\boldsymbol{\xi}, z) - \bar{u}\big\|_U\Big] \ \text{ subject to } \ \alpha \le \mathbb{E}\Big[g_1\big(s(\boldsymbol{\xi}, z), z\big)\Big] \le \beta,$$

which is of the form (2.1) with $m = 2$, $h(w) = \iota_{[\alpha,\beta]}(w_1) + w_2$, $f(z) = 0$, and $g_2(u, z) = \|u - \bar{u}\|_U$. The bounds $\alpha \le \beta$ could coincide to produce an equality constraint or they might have $\alpha = -\infty$ and $\beta = 0$ to produce the inequality $\mathbb{E}[g_1(s(\boldsymbol{\xi}, z), z)] \le 0$.

Example 2.3 (risk modeling). Modeling focused on "worst-case" performance instead of "average" performance is accomplished using a regular measure of risk that maps random variables with finite second moments into $\overline{\mathbb{R}}$ [31, 38, Chapter 8]. Let $\mathcal{R}_i$, $i = 1, \ldots, m$, be a collection of regular measures of risk. Suppose that we seek a control $\hat{z} \in \hat{A} \subset \hat{Z}$ that minimizes the risk associated with a performance function $\hat{g}_1 : U \times \hat{Z} \to \mathbb{R}$ and satisfies constraints of nonnegative risk related to $\hat{g}_i : U \times \hat{Z} \to \mathbb{R}$, $i = 2, 3, \ldots, m$. This leads to the problem

$$\underset{\hat{z} \in \hat{A}}{\text{minimize}} \ \mathcal{R}_1\Big(\hat{g}_1\big(s(\boldsymbol{\xi}, \hat{z}), \hat{z}\big)\Big) \ \text{ subject to } \ \mathcal{R}_i\Big(\hat{g}_i\big(s(\boldsymbol{\xi}, \hat{z}), \hat{z}\big)\Big) \le 0, \ i = 2, 3, \ldots, m,$$

which turns out to be expressible in the form (2.1) for common measures of risk.

Details. By [38, Thm. 8.9], every regular measure of risk $\mathcal{R}$ can be expressed as $\mathcal{R}(\boldsymbol{\eta}) = \min_{\gamma \in \mathbb{R}} \gamma + \mathcal{V}(\boldsymbol{\eta} - \gamma)$ for some regular measure of regret $\mathcal{V}$, which also maps random variables with finite second moments into $\overline{\mathbb{R}}$. Thus, the problem is equivalently stated as a minimization problem over $\hat{z}$ and auxiliary variables $\gamma_1, \ldots, \gamma_m$ using regular measures of regret $\mathcal{V}_1, \ldots, \mathcal{V}_m$ corresponding to $\mathcal{R}_1, \ldots, \mathcal{R}_m$, respectively. Many common measures of regret $\mathcal{V}$ are of the expectation kind, which means that for some convex function $v : \mathbb{R} \to \mathbb{R}$ one has $\mathcal{V}(\boldsymbol{\eta}) = \mathbb{E}[v(\boldsymbol{\eta})]$ for random variables $\boldsymbol{\eta}$; see [33] and [38, Chapter 8]. For $\alpha \in (0, 1)$, the *penalty regret* $\mathcal{V}(\boldsymbol{\eta}) = \mathbb{E}[\max\{0, \boldsymbol{\eta}\}]/(1 - \alpha)$ is a prominent example producing the superquantile risk (a.k.a. CVaR and AVaR) [32]. If $\mathcal{V}_1, \ldots, \mathcal{V}_m$ are of the expectation kind and expressible using the convex functions $v_1, \ldots, v_m$, then the problem takes the equivalent form

$$\operatorname*{minimize}_{\hat{z}\in\hat{A},\gamma_1,\ldots,\gamma_m} \mathbb{E}\Big[\gamma_1 + v_1\big(\hat{g}_1\big(s(\boldsymbol{\xi},\hat{z}),\hat{z}\big) - \gamma_1\big)\Big]$$

$$\text{subject to} \ \ \mathbb{E}\Big[\gamma_i + v_i\big(\hat{g}_i\big(s(\boldsymbol{\xi},\hat{z}),\hat{z}\big) - \gamma_i\big)\Big] \le 0, \ \ i = 2,3,\ldots,m,$$

which indeed is of the form (2.1) with control space $Z = \hat{Z} \times \mathbb{R}^m$, set of admissible controls $A = \hat{A} \times \mathbb{R}^m$, performance functions

$$g_i(u,z) = \gamma_i + v_i\big(\hat{g}_i(u,\hat{z}) - \gamma_i\big), \, i = 1,\ldots,m \ \ \text{for} \ \ z = (\hat{z},\gamma_1,\ldots,\gamma_m),$$

and monitoring function $h(w) = w_1 + \iota_{(-\infty,0]}(w_2) + \cdots + \iota_{(-\infty,0]}(w_m)$.

Example 2.4 (buffered failure probability constraint). A reliability constraint can beneficially be expressed in terms of a buffered failure probability [34], which is better behaved mathematically than a failure probability [25, 5]. For a random variable $\boldsymbol{\eta}$ with finite mean, the buffered failure probability is defined as

$$\text{b-prob}\{\boldsymbol{\eta} > 0\} = \begin{cases} 0 & \text{if} \ \ \text{prob}\{\boldsymbol{\eta} > 0\} = 0, \\ 1 - \alpha & \text{if} \ \ \text{prob}\{\boldsymbol{\eta} > 0\} > 0 \ \ \text{and} \ \ \mathbb{E}[\boldsymbol{\eta}] < 0, \\ 1 & \text{otherwise}, \end{cases}$$

where $\alpha \in (0,1)$ is the probability that makes the α-superquantile $\bar{Q}_\alpha(\boldsymbol{\eta})$ of $\boldsymbol{\eta}$ equal to zero. As expressed by the α-quantile $Q_\alpha(\boldsymbol{\eta})$ of $\boldsymbol{\eta}$, the α-superquantile is defined as [38, section 3.C]

$$\bar{Q}_\alpha(\boldsymbol{\eta}) = Q_\alpha(\boldsymbol{\eta}) + \frac{1}{1 - \alpha}\mathbb{E}\Big[\max\big\{0, \boldsymbol{\eta} - Q_\alpha(\boldsymbol{\eta})\big\}\Big].$$

The problem of determining a control $\hat{z} \in \hat{A} \subset \hat{Z}$ that minimizes a cost function $\hat{f} : \hat{Z} \to \mathbb{R}$ plus the average value of a quantity of interest represented by $\hat{g}_1 : U \times \hat{Z} \to \mathbb{R}$ and produces sufficient reliability relative to another quantity of interest expressed by $\hat{g}_2 : U \times \hat{Z} \to \mathbb{R}$ is formulated as

(2.3)
$$\operatorname*{minimize}_{\hat{z}\in\hat{A}} \hat{f}(\hat{z}) + \mathbb{E}\Big[\hat{g}_1\big(s(\boldsymbol{\xi},\hat{z}),\hat{z}\big)\Big]$$

$$\text{subject to b-prob}\Big\{\hat{g}_2\big(s(\boldsymbol{\xi},\hat{z}),\hat{z}\big) > 0\Big\} \le 1 - \alpha.$$

The problem can be stated in the form (2.1).

Details. For $\alpha \in (0,1)$, $\bar{Q}_\alpha(\boldsymbol{\eta}) \le 0$ if and only if b-prob$\{\boldsymbol{\eta} > 0\} \le 1 - \alpha$ [38, section 3.E]. This fact together with the discussion in Example 2.3 about the penalty regret imply that (2.3) is equivalent to

(2.4)
$$\operatorname*{minimize}_{\hat{z}\in\hat{A},\gamma\in\mathbb{R}} \hat{f}(\hat{z}) + \mathbb{E}\Big[\hat{g}_1\big(s(\boldsymbol{\xi},\hat{z}),\hat{z}\big)\Big]$$

$$\text{subject to} \ \ \mathbb{E}\Big[\gamma + \frac{1}{1 - \alpha}\max\big\{0, \hat{g}_2\big(s(\boldsymbol{\xi},\hat{z}),\hat{z}\big) - \gamma\big\}\Big] \le 0.$$

This problem is of the form (2.1) with $Z = \hat{Z} \times \mathbb{R}$, $A = \hat{A} \times \mathbb{R}$, $h(w_1,w_2) = w_1 + \iota_{(-\infty,0]}(w_2)$, and, for $z = (\hat{z},\gamma)$, $f(z) = \hat{f}(\hat{z})$, $g_1(u,z) = \hat{g}_1(u,\hat{z})$, and $g_2(u,z) = \gamma + \max\{0, \hat{g}_2(u,\hat{z}) - \gamma\}/(1 - \alpha)$.

2.2. Approximations. The examples illustrate the breadth of situations addressed by (2.1) but also the significant challenges associated with the development and justification of computational procedures. In addition to the need for discretization of the control space Z and numerical solution of the underlying PDE, m expectations must be estimated, and potentially nonsmooth performance functions might emerge from reformulations of risk measures (see Examples 2.3 and 2.4) causing the need for smoothing. Nonlinear constraints commonly require approximations via augmented Lagrangian and penalty formulations. These considerations produce approximations of (2.1), where Z is replaced by a finite-dimensional subspace, the monitoring function h is replaced by smooth alternatives, expectations are replaced by sample average approximations, and the quantities of interest, as summarized by G, are replaced by approximations capturing numerical solutions of the underlying PDE, and also smooth approximations necessitated by nonsmooth performance functions.

In the setting of Example 2.4, one might face five types of approximations: A finite-dimensional subspace $\hat{Z}^n$ restricts $\hat{Z}$, a sample $\{\xi_1, \ldots, \xi_\nu\}$ furnishes an estimate of expectations, an approximating solution mapping s^ν replaces s, an augmented Lagrangian term with parameter $\theta^\nu \in [0, \infty)$ and multiplier $y^\nu \in \mathbb{R}$ substitutes for the constraint, and $\mathrm{smax}(\gamma; \beta^\nu) = \beta^\nu \ln(1 + \exp(\gamma/\beta^\nu))$ approximates the max-function $\gamma \mapsto \max\{0, \gamma\}$ with an error of at most $2\beta^\nu$ using a tunable parameter $\beta^\nu \in (0, \infty)$; see [38, Ex. 4.16].

Before implementing these approximations, we reformulate (2.4) using a nonnegative slack variable σ:

$$
(2.5) \quad \begin{aligned}
&\underset{\hat{z} \in \hat{A}, \gamma \in \mathbb{R}, \sigma \geq 0}{\mathrm{minimize}} \; \hat{f}(\hat{z}) + \mathbb{E}\Big[\hat{g}_1\big(s(\boldsymbol{\xi}, \hat{z}), \hat{z}\big)\Big] \\
&\text{subject to} \quad \mathbb{E}\Big[\sigma + \gamma + \frac{1}{1-\alpha} \max\Big\{0, \hat{g}_2\big(s(\boldsymbol{\xi}, \hat{z}), \hat{z}\big) - \gamma\Big\}\Big] = 0.
\end{aligned}
$$

The problem is of the form (2.1) with $Z = \hat{Z} \times \mathbb{R}^2$, $A = \hat{A} \times \mathbb{R} \times [0, \infty)$, $h(w_1, w_2) = w_1 + \iota_{\{0\}}(w_2)$, and, for $z = (\hat{z}, \gamma, \sigma)$, one has $f(z) = \hat{f}(\hat{z})$, $g_1(u, z) = \hat{g}_1(u, \hat{z})$ and

$$
(2.6) \quad g_2(u, z) = \sigma + \gamma + \frac{1}{1-\alpha} \max\big\{0, \hat{g}_2(u, \hat{z}) - \gamma\big\}.
$$

An approximating problem for (2.5) is then

$$
(2.7) \quad \begin{aligned}
&\underset{\hat{z} \in \hat{A} \cap \hat{Z}^n, \gamma \in \mathbb{R}, \sigma \geq 0}{\mathrm{minimize}} \; \hat{f}(\hat{z}) + \frac{1}{\nu} \sum_{j=1}^{\nu} \hat{g}_1\big(s^\nu(\xi_j, \hat{z}), \hat{z}\big) \\
&\quad + \frac{y^\nu}{\nu} \sum_{j=1}^{\nu} \sigma + \gamma + \frac{1}{1-\alpha} \mathrm{smax}\Big(\hat{g}_2\big(s^\nu(\xi_j, \hat{z}), \hat{z}\big) - \gamma; \beta^\nu\Big) \\
&\quad + \theta^\nu \left(\frac{1}{\nu} \sum_{j=1}^{\nu} \sigma + \gamma + \frac{1}{1-\alpha} \mathrm{smax}\Big(\hat{g}_2\big(s^\nu(\xi_j, \hat{z}), \hat{z}\big) - \gamma; \beta^\nu\Big)\right)^2,
\end{aligned}
$$

where the objective function being minimized is an augmented Lagrangian [38, Ex. 6.10]. Presumably, the approximating problem does not involve any difficult constraints, has a smooth objective function in many practical settings, and can be addressed using standard nonlinear programming solvers.

One could hope that solutions of the approximating problem produce reasonable solutions for (2.3) when the subspace $\hat{Z}^n$ is sufficiently close to $\hat{Z}$, the sample size ν

is large, the numerical solution mapping s^ν approximates s rather well, the penalty parameter θ^ν is high, the multiplier y^ν is suitably selected, and the smoothing parameter β^ν is near zero. Of course, this cannot be expected in general. The next sections examine performance bounds for such approximations in the concrete setting of (2.5) and also more broadly.

3. General approximation theory. To prepare a foundation for a unified treatment of the actual problem (2.1), we approach an abstract optimization problem and its approximations. Suppose that (X, d_X) is a metric space[1] and let $\mathbb{N} = \{1, 2, \dots\}$. For $f : X \to \mathbb{R}$, consider the problem

$$(3.1) \qquad \qquad \underset{x \in X}{\text{minimize}}\, f(x)$$

and the discretized problems

$$(3.2) \qquad \left\{ \underset{x \in X}{\text{minimize}}\, f(x) + \iota_{X^n}(x),\ n \in \mathbb{N} \right\},$$

where $X^n \subset X$ corresponds to a finite-dimensional space. Specifically, for each $n \in \mathbb{N}$, we assume there is a surjective mapping $T_n : \mathbb{R}^n \to X^n$, i.e., for every $x \in X^n$, one can identify a vector $x_n \in \mathbb{R}^n$ satisfying $x = T_n(x_n)$. Through this correspondence, we identify the finite-dimensional problems

$$(3.3) \qquad \left\{ \underset{x_n \in \mathbb{R}^n}{\text{minimize}}\, f_n(x_n),\ n \in \mathbb{N} \right\},$$

where $f_n : \mathbb{R}^n \to \overline{\mathbb{R}}$ is given by

$$(3.4) \qquad \qquad f_n(x_n) = f\big(T_n(x_n)\big).$$

We see that (3.2) and (3.3) are equivalent, with

$$(3.5) \qquad \inf\big\{ f_n(x_n) \,\big|\, x_n \in \mathbb{R}^n \big\} = \inf\big\{ f(x) + \iota_{X^n}(x) \,\big|\, x \in X \big\}.$$

Although finite dimensional, (3.3) might require approximations for computational and other reasons. We consider the collection of approximating problems

$$(3.6) \qquad \forall n \in \mathbb{N}:\ \left\{ \underset{x_n \in \mathbb{R}^n}{\text{minimize}}\, f_n^\nu(x_n),\ \nu \in \mathbb{N} \right\},$$

where $f_n^\nu : \mathbb{R}^n \to \overline{\mathbb{R}}$ is an approximation of f_n that presumably becomes more accurate as $\nu \to \infty$. We seek to justify the following approach to solving (3.1): apply an optimization algorithm to one or more of the approximating problems (3.6) and obtain a solution that serves as an approximating solution of (3.1). As a preview of our treatment of the actual problem (2.1), the finite-dimensional problems (3.3) correspond to a discretization of the control space and the approximating problems (3.6) implement all other approximations as well, including discretization adopted to solve the underlying PDE. Since discretization of the control space can be viewed as just another approximation, it appears somewhat arbitrary to single out that discretization. However, this separation is technically convenient and allows us to utilize the fact that discretization of the control space imposes a restriction on the problem while the other approximations may not have that characteristic.

---

[1] The metric space setting here hints at the possibility of extending the reach of later results beyond Banach spaces.

We adopt the following notation and concepts. For $\delta \in [0, \infty)$ and a function $g : X \to \overline{\mathbb{R}}$, the set of near-minimizers is given by

$$\delta\text{-argmin}\, g = \big\{ x \in \text{dom}\, g \mid g(x) \leq \inf g + \delta \big\},$$

where $\inf g = \inf \{ g(x) \mid x \in X \}$ and $\text{dom}\, g = \{ x \in X \mid g(x) < \infty \}$. Convergence x^ν to x along a subsequence indexed by $N \subset \mathbb{N}$ is specified by $x^\nu \underset{N}{\to} x$. The index set N is then taken from the collection of subsequences of $\mathbb{N}$ denoted by $\mathcal{N}_\infty^\#$.

DEFINITION 3.1 (epi-convergence). *For $g^\nu, g : X \to \overline{\mathbb{R}}$, we say that g^ν epi-converges to g as $\nu \to \infty$, written $g^\nu \overset{e}{\to} g$, when the following hold at every $x \in X$:*
  (a) *For all $x^\nu \to x$, one has $\liminf g^\nu(x^\nu) \geq g(x)$.*
  (b) *There exists $x^\nu \to x$ such that $\limsup g^\nu(x^\nu) \leq g(x)$.*

The main result of this section furnishes an optimality gap for a control obtained by solving the approximating problems (3.6).

THEOREM 3.2 (optimality gap). *For a metric space (X, d_X), suppose that $f : X \to \overline{\mathbb{R}}$, $X^n \subset X$, and $T_n : \mathbb{R}^n \to X^n$ is a surjective mapping. Let $f_n, f_n^\nu : \mathbb{R}^n \to \overline{\mathbb{R}}$, with f_n defined by (3.4). Suppose that the following hold for some $\bar{n} \in \mathbb{N}$:*
  (a) *$f + \iota_{X^n} \overset{e}{\to} f$ as $n \to \infty$.*
  (b) *$\forall n \geq \bar{n},\ f_n^\nu \overset{e}{\to} f_n$ as $\nu \to \infty$.*
  (c) *$\inf f > -\infty$ and, $\forall n \geq \bar{n},\ \inf f_n < \infty$.*
Then, for any $\varepsilon \in (0, \infty)$, there exists $n_\varepsilon \geq \bar{n}$ such that when $n \geq n_\varepsilon$, $\delta^\nu \to \delta \in [0, \infty)$, and $\{ \bar{x}_n^\nu \in \delta^\nu\text{-}\text{argmin}\, f_n^\nu, \nu \in \mathbb{N} \}$ has a cluster point $\bar{x}_n$, one obtains

$$f\big(T_n(\bar{x}_n)\big) \leq \inf f + \varepsilon + \delta.$$

Proof. Suppose that $\inf f$ is finite and let $\gamma \in (0, \infty)$. Then, there is $x \in X$ such that $f(x) \leq \inf f + \gamma$. In view of (a) and Definition 3.1(b), there is $x^n \to x$ such that $\limsup_n (f(x^n) + \iota_{X^n}(x^n)) \leq f(x)$. Stringing these two inequalities together, we obtain that

$$\limsup_n \big(\inf(f + \iota_{X^n}) \big) \leq \limsup_n \big(f(x^n) + \iota_{X^n}(x^n) \big) \leq f(x) \leq \inf f + \gamma.$$

Since γ is arbitrary, we conclude that

$$(3.7) \qquad\qquad \limsup_n \big(\inf(f + \iota_{X^n}) \big) \leq \inf f$$

when $\inf f$ is finite. The same holds trivially when $\inf f = \infty$. The possibility $\inf f = -\infty$ is ruled out by (c).

Let $\varepsilon \in (0, \infty)$. By (c) and (3.7), $\inf f > -\infty$ and there is $n_\varepsilon \geq \bar{n}$ such that

$$(3.8) \qquad\qquad \inf(f + \iota_{X^n}) \leq \inf f + \varepsilon \quad \forall n \geq n_\varepsilon.$$

Fix $n \geq n_\varepsilon$. First, we establish that

$$(3.9) \qquad\qquad \limsup_\nu (\inf f_n^\nu) \leq \inf f_n.$$

An argument parallel to the one carried out to reach (3.7) confirms the claim when $\inf f_n$ is finite. Since $\inf f_n = \infty$ is ruled out by (c), it only remains to examine the case with $\inf f_n = -\infty$. Let $\gamma \in (0, \infty)$. Then, there is $x_n \in \mathbb{R}^n$ such that $f_n(x_n) \leq -\gamma$ and also, by Definition 3.1(b), there are $x_n^\nu \to x_n$ such that $\limsup_\nu f_n^\nu(x_n^\nu) \leq f_n(x_n)$.

Thus,

$$\text{limsup}_\nu(\inf f_n^\nu) \leq \text{limsup}_\nu f_n^\nu(x_n^\nu) \leq f_n(x_n) \leq -\gamma.$$

Since γ is arbitrary, (3.9) holds in this case as well.

Second, let $N \in \mathcal{N}_\infty^\#$ be the subsequence corresponding to the cluster point $\bar{x}_n$, i.e., $\bar{x}_n^\nu \xrightarrow[N]{} \bar{x}_n$. Then, (3.9) and the construction of $\{\bar{x}_n^\nu, \nu \in \mathbb{N}\}$ ensure that

$$\text{limsup}_{\nu \in N} f_n^\nu(\bar{x}_n^\nu) \leq \text{limsup}_{\nu \in N}(\inf f_n^\nu + \delta^\nu) \leq \inf f_n + \delta.$$

By Definition 3.1(a), this implies that

$$f_n(\bar{x}_n) \leq \text{liminf}_{\nu \in N} f_n^\nu(\bar{x}_n^\nu) \leq \text{limsup}_{\nu \in N} f_n^\nu(\bar{x}_n^\nu) \leq \inf f_n + \delta.$$

Since $\inf f_n < \infty$ by (c), $\bar{x}_n \in \text{dom} f_n$. Thus, $\bar{x}_n \in \delta\text{-argmin} f_n$. We then obtain

$$f\big(T_n(\bar{x}_n)\big) = f_n(\bar{x}_n) \leq \inf f_n + \delta = \inf(f + \iota_{X^n}) + \delta \leq \inf f + \varepsilon + \delta,$$

where the second equality follows by (3.5) and the second inequality by (3.8). □

Theorem 3.2 provides an optimality gap, relative to the infinite-dimensional problem (3.1), for any cluster point constructed by solving a sequence of approximating finite-dimensional problems. The gap is bounded by the sum of two terms: ε and δ. While ε can be selected arbitrarily near zero, this typically results in a large n_ε and thus the need for minimizing a high-dimensional function f_n^ν with a tolerance eventually near δ. The theorem applies for any δ so the approximating problems do not need to be solved to a high tolerance. The tolerance might not even be fully known. Still, the theorem provides a guarantee that better solutions of the approximating problems (i.e., lower δ^ν) translate into better solutions of the infinite-dimensional problem (3.1), at least in the sense of an upper bound.

The theorem does not require convexity, smoothness, or continuity of the functions f, f_n, and f_n^ν. However, the functions f and f_n are lower semicontinuous (lsc) by virtue of being epi-limits. We note that $\{\bar{x}_n^\nu, \nu \in \mathbb{N}\}$ is a sequence in $\mathbb{R}^n$ and thus has a cluster point as long as the sequence is bounded.

In general, ε can be thought of as a *discretization error* due to the restriction from X to X^n, and δ^ν (and its limiting value δ) is an *optimization error* caused by our incomplete minimization of f_n^ν. There is no *approximation error* in the theorem, something one might expect from replacing f_n in (3.3) by f_n^ν in (3.6). This is avoided because, as $\nu \to \infty$, such errors vanish by assumption (b) in Theorem 3.2.

To the best of our knowledge, there is no prior result matching Theorem 3.2. The novelty lies in the consideration of doubly indexed approximating functions f_n^ν on $\mathbb{R}^n$ that do not tend to the actual function f but rather to f_n. In fact, f_n^ν and f are not even defined on the same space so any talk of their epi-convergence is not meaningful. As we will see, Theorem 3.2 emerges as a key building block in the next section but can also stand on its own as a valuable fact for general optimization problems on metric spaces.

4. Performance bounds. We now return to PDE-constrained optimization under uncertainty (OUU) and the notation in section 2. Following the pattern of section 3 with (3.1) being replaced by the problems (3.2), (3.3), and (3.6), we start by defining the *control-discretized problems*

$$(4.1) \qquad \Big\{ \underset{z \in Z}{\text{minimize}}\ \varphi(z) + \iota_{Z^n}(z),\ n \in \mathbb{N} \Big\},$$

where $\varphi : Z \to \overline{\mathbb{R}}$ is the objective function in the actual problem (2.1) and $Z^n \subset Z$. We assume that elements in Z^n can be represented by n real-valued parameters and that there is a surjective mapping $T_n : \mathbb{R}^n \to Z^n$. Through this correspondence, we obtain the *equivalent control-discretized problems*

$$(4.2) \qquad \Big\{ \underset{z_n \in \mathbb{R}^n}{\text{minimize}} \; \varphi_n(z_n), \; n \in \mathbb{N} \Big\},$$

where $\varphi_n : \mathbb{R}^n \to \overline{\mathbb{R}}$ is given by $\varphi_n(z_n) = \varphi(T_n(z_n))$. Leveraging the definition of φ in (2.1), the problems in (4.2) take the form

$$(4.3) \qquad \underset{z_n \in \mathbb{R}^n}{\text{minimize}} \; \varphi_n(z_n) = \iota_{A_n}(z_n) + f_n(z_n) + h\Big(\mathbb{E}\big[G_n(\boldsymbol{\xi}, z_n) \big] \Big),$$

where $A_n = \{ z_n \in \mathbb{R}^n \mid T_n(z_n) \in A \}$, $f_n(z_n) = f(T_n(z_n))$, and $G_n(\xi, z_n) = G(\xi, T_n(z_n))$ so that $f_n : \mathbb{R}^n \to \mathbb{R}$ and $G_n : \Xi \times \mathbb{R}^n \to \mathbb{R}^m$.

Assumption 4.1 (continuity). For a.e. $\xi \in \Xi$ and $i = 1, \ldots, m$, $s(\xi, \cdot) : Z \to U$, g_i, and T_n are continuous.[2] For every $z_n \in \mathbb{R}^n$, $G_n(\cdot, z_n) : \Xi \to \mathbb{R}^m$ is measurable.

Under Assumption 4.1, $\mathbb{E}[G_n(\boldsymbol{\xi}, z_n)]$ is well defined for each $z_n \in \mathbb{R}^n$ under the usual convention of setting an expectation to infinity if both the positive part and the negative part integrate to infinity [38, section 3.B].

Next, we turn to approximations of the equivalent control-discretized problems. For $\xi_1, \ldots, \xi_\nu \in \Xi$, the *approximating problems* take the form

$$(4.4) \qquad \underset{z_n \in \mathbb{R}^n}{\text{minimize}} \; \varphi_n^\nu(z_n) = \iota_{A_n}(z_n) + f_n(z_n) + h^\nu \left(\frac{1}{\nu} \sum_{j=1}^{\nu} G_n^\nu(\xi_j, z_n) \right),$$

where a sample average replaces the expectation, $h^\nu : \mathbb{R}^m \to \overline{\mathbb{R}}$ approximates the monitoring function, and $G_n^\nu : \Xi \times \mathbb{R}^n \to \mathbb{R}^m$ approximates G_n and is given by

$$(4.5) \qquad G_n^\nu(\xi, z_n) = G^\nu\big(\xi, T_n(z_n) \big).$$

Here, $G^\nu : \Xi \times Z \to \mathbb{R}^m$ has

$$(4.6) \qquad G^\nu(\xi, z) = \Big(g_1^\nu\big(s^\nu(\xi, z), z \big), \ldots, g_m^\nu\big(s^\nu(\xi, z), z \big) \Big),$$

where $g_i^\nu : U \times Z \to \mathbb{R}$ approximates g_i and $s^\nu : \Xi \times Z \to U$ approximates $s : \Xi \times Z \to U$; see section 5. We consider the following scheme to solve (2.1).

Approximation Algorithm for PDE-Constrained OUU.

Data. $n \in \mathbb{N}$ and $\{ \xi_j \in \Xi, j \in \mathbb{N} \}$.

Step 0. Set $\nu = 1$.

Step 1. Apply an algorithm to (4.4), obtain $\bar{z}_n^\nu \in \mathbb{R}^n$, and record $\varphi_n^\nu(\bar{z}_n^\nu)$.

Step 2. Replace ν by $\nu + 1$ and go to Step 1.

Presumably, the approximating problems (4.4) are tractable for existing (nonlinear programming) algorithms, but there is no requirement that $\bar{z}_n^\nu$ must be globally or locally optimal or even stationary for φ_n^ν. A main feature of the subsequent analysis, already alluded to in section 3, is that the resulting performance guarantee holds even though Step 1 is not carried out "perfectly." In particular, if $\{ \bar{z}_n^\nu, \nu \in \mathbb{N} \}$ has a cluster point $\bar{z}_n$, which indeed would be the case when the sequence is bounded, then the

---

[2]Throughout, a product space is equipped with the product norm.

recorded values $\{\varphi_n^\nu(\bar{z}_n^\nu), \nu \in \mathbb{N}\}$ lead to a performance guarantee for $\bar{z}_n$ as measured by the actual objective function φ; see Theorem 4.9 below.

Numerous implementation issues emerge including rules for refining approximation levels and stopping criteria in Step 1. It would typically be inefficient to apply much effort in Step 1 for *each* ν. In practice, one might leapfrog over most ν and invest computing resources towards, say, $\nu = 1000, 2000, 3000$, etc. Each of the expectations might also demand different sample sizes. They are all set to ν in this paper for notational convenience, but the results extend to other schemes trivially. While n is kept fixed in the algorithm, an implementation might also involve a gradual increase of that parameter.

4.1. Intermediate results. This subsection presents assumptions and furnishes several technical results. In the following, inequalities between vectors are assumed to apply componentwise. The Euclidean balls in $\mathbb{R}^n$ are written as

$$\mathbb{B}(\bar{z}_n, \rho) = \{z_n \in \mathbb{R}^n \mid \|z_n - \bar{z}_n\|_2 \le \rho\}.$$

Assumption 4.2 (locally bounded quantities of interest). For each $\bar{z}_n \in A_n$, there are $\rho \in (0, \infty)$ and P-integrable $c : \Xi \to [0, \infty)$ such that

$$\big\|G_n(\xi, z_n)\big\|_\infty \le c(\xi) \ \ \forall z_n \in \mathbb{B}(\bar{z}_n, \rho) \cap A_n, \ \text{a.e. } \xi \in \Xi.$$

LEMMA 4.3. *Let $n \in \mathbb{N}$ be fixed. Suppose that A_n is closed, Assumptions 4.1 and 4.2 hold for n, and the sample $\{\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \dots\}$ is independent and identically distributed (i.i.d.) as $\boldsymbol{\xi}$. Then, with probability one,*

$$\lim_\nu \frac{1}{\nu} \sum_{j=1}^\nu G_n(\boldsymbol{\xi}_j, z_n^\nu) = \mathbb{E}\big[G_n(\boldsymbol{\xi}, z_n)\big] \in \mathbb{R}^m \ \ \text{whenever } z_n^\nu \in A_n \to z_n.$$

Proof. Let $\psi_i : \Xi \times \mathbb{R}^n \to \mathbb{R}$ be given by $\psi_i(\xi, z_n) = g_i(s(\xi, T_n(z_n)), T_n(z_n))$. For a.e. $\xi \in \Xi$, $\psi_i(\xi, \cdot)$ is continuous by Assumption 4.1. For all $z_n \in \mathbb{R}^n$, $\psi_i(\cdot, z_n)$ is measurable. Thus, ψ_i is a Caratheodory function and certainly random lsc; see [38, Ex. 8.51(d)]. Moreover, $(\xi, z_n) \mapsto \psi_i(\xi, z_n) + \iota_{A_n}(z_n)$ is random lsc by [38, Ex. 8.51(c)] because A_n is closed. With probability one, the epigraphical law of large numbers [38, Thm. 8.56] applied to each ψ_i yields

$$\liminf_\nu \frac{1}{\nu} \sum_{j=1}^\nu G_n(\boldsymbol{\xi}_j, z_n^\nu) \ge \mathbb{E}\big[G_n(\boldsymbol{\xi}, z_n)\big] \ \ \text{whenever } z_n^\nu \in A_n \to z_n,$$

with no component of $\mathbb{E}[G_n(\boldsymbol{\xi}, z_n)]$ being $-\infty$. Repeating this argument with $-\psi_i$ in place of ψ_i, we obtain the conclusion. $\qquad\square$

Assumption 4.4 (approximation of performance functions). There is a sequence $\{\varepsilon^\nu \in [0, \infty), \nu \in \mathbb{N}\}$ with $\varepsilon^\nu \to 0$ such that

$$\big|g_i^\nu(u, z) - g_i(u, z)\big| \le \varepsilon^\nu \ \ \forall u \in U, z \in Z^n \cap A, i = 1, \dots, m.$$

The assumption requires a uniform error bound for the approximating performance functions, which is satisfied, for example, by the smax-approximation in subsection 2.2. The uniformity can be relaxed, with slight adjustments to Assumption 4.5 below. We omit these details due to the resulting notational complexity. Anyhow, they are not needed for the application in section 5.

Assumption 4.5 (solution properties and approximations). There exist $\eta, \kappa \in (0, \infty)$, $\lambda, \mu : \Xi \to [0, \infty)$, $\rho, \pi : A_n \to (0, \infty)$, and $\{\Delta_0, \Delta_\nu : A_n \to [0, \infty), \nu \in \mathbb{N}\}$ satisfying the following properties:

(a) $\mathbb{E}[\lambda^\eta(\boldsymbol{\xi})\mu^\kappa(\boldsymbol{\xi})]$ and $\mathbb{E}[\mu^\kappa(\boldsymbol{\xi})]$ are finite.

(b) $\forall z_n \in A_n$, $z^n = T_n(z_n)$, $u, u' \in U$, $\nu \in \mathbb{N}$, $i = 1, \ldots, m$, and a.e. $\xi \in \Xi$, one has

$$\left|g_i(u, z^n) - g_i(u', z^n)\right| \le \pi(z_n)\left(\|u\|_U^\eta + \|u'\|_U^\eta + 1\right)\|u - u'\|_U^\kappa,$$

$$\max\left\{\left\|s(\xi, z^n)\right\|_U, \ \left\|s^\nu(\xi, z^n)\right\|_U\right\} \le \lambda(\xi)\Delta_0(z_n),$$

$$\left\|s^\nu(\xi, z^n) - s(\xi, z^n)\right\|_U \le \mu(\xi)\Delta_\nu(z_n).$$

(c) $\forall \bar{z}_n \in A_n$, one has

$$\sup_{z_n \in \mathbb{B}(\bar{z}_n, \rho(\bar{z}_n)) \cap A_n} \pi(z_n) < \infty, \qquad \sup_{z_n \in \mathbb{B}(\bar{z}_n, \rho(\bar{z}_n)) \cap A_n} \Delta_0(z_n) < \infty,$$

$$\sup_{z_n \in \mathbb{B}(\bar{z}_n, \rho(\bar{z}_n)) \cap A_n} \Delta_\nu(z_n) \to 0, \quad \text{as } \nu \to \infty.$$

Part (b) imposes bounds on approximation errors of the solutions as well as a relatively mild Hölder-type condition on the performance functions, which can be made more general depending on specific applications. Section 5 furnishes a specific example of when these assumptions hold.

LEMMA 4.6. *Let $n \in \mathbb{N}$ be fixed. Suppose that A_n is closed, Assumptions 4.4 and 4.5 hold for n, and the sample $\{\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots\}$ is i.i.d. at $\boldsymbol{\xi}$. Then, with probability one,*

$$\frac{1}{\nu}\sum_{j=1}^{\nu}\left\|G_n^\nu(\boldsymbol{\xi}_j, z_n^\nu) - G_n(\boldsymbol{\xi}_j, z_n^\nu)\right\|_\infty \to 0 \quad \text{whenever } z_n^\nu \in A_n \to z_n.$$

Proof. Let $\eta, \kappa, \lambda, \mu, \rho, \pi, \Delta_0, \Delta_\nu$, and ε^ν be as specified by Assumptions 4.4 and 4.5. First, we fix $z_n \in A_n$, $z^n = T_n(z_n)$, $\xi \in \Xi_1$, $\nu \in \mathbb{N}$, and $i \in \{1, \ldots, m\}$, where $P(\Xi_1) = 1$. Trivially,

$$\left|g_i^\nu\big(s^\nu(\xi, z^n), z^n\big) - g_i\big(s(\xi, z^n), z^n\big)\right|$$

$$\le \left|g_i\big(s^\nu(\xi, z^n), z^n\big) - g_i\big(s(\xi, z^n), z^n\big)\right| + \left|g_i^\nu\big(s^\nu(\xi, z^n), z^n\big) - g_i\big(s^\nu(\xi, z^n), z^n\big)\right|.$$

Assumption 4.5 addresses the first term so that

$$\left|g_i\big(s^\nu(\xi, z^n), z^n\big) - g_i\big(s(\xi, z^n), z^n\big)\right|$$

$$\le \pi(z_n)\left(\left\|s^\nu(\xi, z^n)\right\|_U^\eta + \left\|s(\xi, z^n)\right\|_U^\eta + 1\right)\left\|s^\nu(\xi, z^n) - s(\xi, z^n)\right\|_U^\kappa$$

$$\le \pi(z_n)\left(2\lambda^\eta(\xi)\Delta_0^\eta(z_n) + 1\right)\mu^\kappa(\xi)\Delta_\nu^\kappa(z_n).$$

Also bringing in Assumption 4.4, we obtain for $\{\xi_j \in \Xi_1, j \in \mathbb{N}\}$ that

$$\frac{1}{\nu}\sum_{j=1}^{\nu}\left\|G_n^\nu(\xi_j, z_n) - G_n(\xi_j, z_n)\right\|_\infty \le \pi(z_n)2\Delta_0^\eta(z_n)\Delta_\nu^\kappa(z_n)\frac{1}{\nu}\sum_{j=1}^{\nu}\lambda^\eta(\xi_j)\mu^\kappa(\xi_j)$$

$$+ \pi(z_n)\Delta_\nu^\kappa(z_n)\frac{1}{\nu}\sum_{j=1}^{\nu}\mu^\kappa(\xi_j) + \varepsilon^\nu.$$

Second, Assumption 4.5(a) and the i.i.d. sampling ensure that

$$\frac{1}{\nu}\sum_{j=1}^{\nu}\lambda^\eta(\boldsymbol{\xi}_j)\mu^\kappa(\boldsymbol{\xi}_j) \to \mathbb{E}\big[\lambda^\eta(\boldsymbol{\xi})\mu^\kappa(\boldsymbol{\xi})\big] \in \mathbb{R}, \quad \frac{1}{\nu}\sum_{j=1}^{\nu}\mu^\kappa(\boldsymbol{\xi}_j) \to \mathbb{E}\big[\mu^\kappa(\boldsymbol{\xi})\big] \in \mathbb{R}$$

with probability one. Let $\xi_1, \xi_2, \ldots$ be a sequence in Ξ_1 for which this convergence holds and let $z_n^\nu \in A_n \to \bar{z}_n$. Then, there exists $\bar{\nu}$ such that $z_n^\nu \in \mathbb{B}(\bar{z}_n, \rho(\bar{z}_n))$ for all $\nu \geq \bar{\nu}$. Let $B = \mathbb{B}(\bar{z}_n, \rho(\bar{z}_n)) \cap A_n$, $\bar{\pi} = \sup_{z_n \in B} \pi(z_n)$, $\bar{\Delta} = \sup_{z_n \in B} \Delta_0(z_n)$, and $\delta_\nu = \sup_{z_n \in B} \Delta_\nu(z_n)$. Thus, for all $\nu \geq \bar{\nu}$,

$$\frac{1}{\nu} \sum_{j=1}^\nu \left\| G_n^\nu(\xi_j, z_n^\nu) - G_n(\xi_j, z_n^\nu) \right\|_\infty \leq \frac{\bar{\pi} 2 \bar{\Delta}^\eta \delta_\nu^\kappa}{\nu} \sum_{j=1}^\nu \lambda^\eta(\xi_j) \mu^\kappa(\xi_j) + \frac{\bar{\pi} \delta_\nu^\kappa}{\nu} \sum_{j=1}^\nu \mu^\kappa(\xi_j) + \varepsilon^\nu.$$

Since $\bar{\pi}, \bar{\Delta} < \infty$ and $\delta_\nu, \varepsilon^\nu \to 0$, the conclusion follows. $\qquad\square$

Assumption 4.7 (locally bounded quantities of interest on Z). For each $\bar{z} \in A$, $G(\cdot, \bar{z})$ is measurable and there are $\rho \in (0, \infty)$ and P-integrable $c : \Xi \to [0, \infty)$ such that

$$\left\| G(\xi, z) \right\|_\infty \leq c(\xi) \quad \text{for } z \in A, \ \|z - \bar{z}\|_Z \leq \rho, \text{ a.e. } \xi \in \Xi.$$

This assumption resembles Assumption 4.2 but is stated separately to avoid intricacies about the relation between Z and $\mathbb{R}^n$.

LEMMA 4.8 (continuity of expectation function). *If A is closed and Assumptions 4.1 and 4.7 hold, then $\mathbb{E}[G(\boldsymbol{\xi}, z^n)] \to \mathbb{E}[G(\boldsymbol{\xi}, z)] \in \mathbb{R}^m$ whenever $z^n \in A \to z$.*

Proof. Under the stated assumptions, $G(\xi, \cdot)$ is continuous for a.e. ξ. Thus, the fact follows from a standard application of the dominated convergence theorem. $\qquad\square$

4.2. Main results. We establish two performance guarantees related to the Approximation Algorithm for PDE-Constrained OUU. The first one furnishes an upper bound on the objective function value in the actual problem (2.1) for any cluster point produced by the algorithm. The second one specifies an optimality gap.

For metric space X, a function $g : X \to \overline{\mathbb{R}}$ is lsc relative to $C \subset X$ if $\liminf g(x^\nu) \geq g(x)$ whenever $x^\nu \in C \to x \in C$.

THEOREM 4.9 (upper bound in PDE-constrained OUU). *For fixed $n \in \mathbb{N}$, suppose that A_n is closed, f_n is lsc relative to A_n, the sample $\{\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots\}$ is i.i.d. as $\boldsymbol{\xi}$, Assumptions 4.1, 4.2, 4.4, and 4.5 hold for n, and*

$$\liminf h^\nu(w^\nu) \geq h(w) > -\infty \quad \text{whenever } w^\nu \to w \in \mathbb{R}^m.$$

With probability one, if $\{\bar{z}_n^\nu \in \mathbb{R}^n, \nu \in \mathbb{N}\}$ converges to $\bar{z}_n$ along a subsequence $N \in \mathcal{N}_\infty^\#$ as $\nu \to \infty$, then

$$\varphi\big(T_n(\bar{z}_n)\big) \leq \liminf_{\nu \in N} \varphi_n^\nu(\bar{z}_n^\nu).$$

Proof. By Lemmas 4.3 and 4.6, with probability one, $z_n^\nu \in A_n \to z_n$ implies

$$(4.7) \qquad \begin{aligned} &\frac{1}{\nu} \sum_{j=1}^\nu G_n(\boldsymbol{\xi}_j, z_n^\nu) \to \mathbb{E}\big[G_n(\boldsymbol{\xi}, z_n)\big] \in \mathbb{R}^m, \\ &\frac{1}{\nu} \sum_{j=1}^\nu \left\| G_n^\nu(\boldsymbol{\xi}_j, z_n^\nu) - G_n(\boldsymbol{\xi}_j, z_n^\nu) \right\|_\infty \to 0. \end{aligned}$$

Let $\{\xi_1, \xi_2, \ldots\}$ be an event for which this occurs. Then,

$$\frac{1}{\nu} \sum_{j=1}^\nu G_n^\nu(\xi_j, z_n^\nu) = \frac{1}{\nu} \sum_{j=1}^\nu \big(G_n^\nu(\xi_j, z_n^\nu) - G_n(\xi_j, z_n^\nu)\big) + \frac{1}{\nu} \sum_{j=1}^\nu G_n(\xi_j, z_n^\nu) \to \mathbb{E}\big[G_n(\boldsymbol{\xi}, z_n)\big]$$

whenever $z_n^\nu \in A_n \to z_n$.

Suppose that $\bar{z}_n^\nu \xrightarrow[N]{} \bar{z}_n$ for $N \in \mathcal{N}_\infty^\#$. We aim to show that

$$(4.8) \qquad \liminf_{\nu \in N} \varphi_n^\nu(\bar{z}_n^\nu) \geq \varphi_n(\bar{z}_n)$$

because then the conclusion follows from the fact that $\varphi_n(\bar{z}_n) = \varphi(T_n(\bar{z}_n))$.

If $\bar{z}_n \notin A_n$, then $\bar{z}_n^\nu \notin A_n$ for sufficiently large $\nu \in N$ because A_n is closed. Thus, (4.8) holds trivially in this case, and we concentrate on the case with $\bar{z}_n \in A_n$. Without loss of generality, we assume that $\bar{z}_n^\nu \in A_n$ for all $\nu \in N$ because any $\bar{z}_n^\nu \notin A_n$ produces $\varphi_n^\nu(\bar{z}_n^\nu) = \infty$. These facts and the assumption on the monitoring functions imply that

$$\liminf_{\nu \in N} h^\nu \left(\frac{1}{\nu} \sum_{j=1}^{\nu} G_n^\nu(\xi_j, \bar{z}_n^\nu) \right) \geq h\Big(\mathbb{E}\big[G_n(\boldsymbol{\xi}, \bar{z}_n) \big] \Big) > -\infty.$$

Since f_n is lsc relative to A_n, one has

$$\liminf_{\nu \in N} f_n(\bar{z}_n^\nu) \geq f_n(\bar{z}_n).$$

All three terms defining $\varphi_n^\nu(\bar{z}_n^\nu)$ are bounded from below. Thus, it follows that

$$\liminf_{\nu \in N} \varphi_n^\nu(\bar{z}_n^\nu)$$

$$\geq \liminf_{\nu \in N} \iota_{A_n}(\bar{z}_n^\nu) + \liminf_{\nu \in N} f_n(\bar{z}_n^\nu) + \liminf_{\nu \in N} h^\nu \left(\frac{1}{\nu} \sum_{j=1}^{\nu} G_n^\nu(\xi_j, \bar{z}_n^\nu) \right)$$

$$\geq \iota_{A_n}(\bar{z}_n) + f_n(\bar{z}_n) + h\Big(\mathbb{E}\big[G_n(\boldsymbol{\xi}, \bar{z}_n) \big] \Big) = \varphi_n(\bar{z}_n),$$

which completes the proof. $\qquad \square$

Theorem 4.9 implies that if the algorithm constructs a sequence $\{\bar{z}_n^\nu, \nu \in \mathbb{N}\}$ with a cluster point, then that point—a finite-dimensional vector—corresponds to a point in the control space Z, which is as good as we would expect from the recorded values $\{\varphi_n^\nu(\bar{z}_n^\nu), \nu \in \mathbb{N}\}$. The theorem holds regardless of n, but a large n would typically be associated with a large set Z^n and thus better chances to obtain low values $\varphi_n^\nu(\bar{z}_n^\nu)$.

The assumption on h^ν and h is satisfied by models where inequality constraints are replaced by penalties. For example, $h(w) = \iota_{(-\infty,0]^m}(w)$ is approximated by $h^\nu(w) = \theta^\nu \sum_{i=1}^m (\max\{0, w_i\})^2$, where $\theta^\nu \to \infty$.

There is no need to solve (4.4) to local or global optimality. One would simply attempt to bring the objective function value down as $\nu \to \infty$. In practice, one might have a goal of obtaining a control $z \in Z$ such that $\varphi(z) \leq \alpha$. The theorem provides a way of certifying this: pick a reasonably large n, and apply the algorithm with a stopping criterion in Step 1 of $\varphi_n^\nu(\bar{z}_n^\nu) \leq \alpha$. Any cluster point $\bar{z}_n$ of the constructed sequence produces a control $\bar{z} = T_n(\bar{z}_n)$, which then is good enough. Since the theorem does not leverage any assumptions on the discretization Z^n of Z, it becomes impossible to certify a priori whether a good solution of this kind is achievable; one needs to wait for the computations and see what values come out.

Next, we introduce additional assumptions, and these allow us to claim that for sufficiently large n, the algorithm obtains a decision that is arbitrarily good relative to the actual problem.

In the following, the ith component of a vector v is sometimes indicated by $(v)_i$.

Assumption 4.10 (constraint qualification). For fixed n and all points $z_n \in A_n$ satisfying $h(\mathbb{E}[G_n(\boldsymbol{\xi}, z_n)]) < \infty$, there exists $z_n^\nu \in A_n \to z_n$ as $\nu \to \infty$ such that

$$\Big(\mathbb{E}\big[G_n(\boldsymbol{\xi}, z_n^\nu) \big] \Big)_i < \Big(\mathbb{E}\big[G_n(\boldsymbol{\xi}, z_n) \big] \Big)_i, \ \ i = 1, \ldots, m, \ \nu \in \mathbb{N}.$$

For all $z \in A$, with $h(\mathbb{E}[G(\boldsymbol{\xi}, z)]) < \infty$, there exists $z^n \in A \cap Z^n \to z$ as $n \to \infty$ such that $\mathbb{E}[G(\boldsymbol{\xi}, z^n)] \leq \mathbb{E}[G(\boldsymbol{\xi}, z)]$ for all $n \in \mathbb{N}$.

THEOREM 4.11 (optimality gap in PDE-constrained OUU). *For fixed $\bar{n} \in \mathbb{N}$, suppose that $\inf \varphi > -\infty$, A is closed, f is continuous relative to A, the sample $\{\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \dots\}$ is i.i.d. as $\boldsymbol{\xi}$, and, for all $n \geq \bar{n}$, $\inf \varphi_n < \infty$, and suppose that Assumptions 4.1, 4.2, 4.4, 4.5, and 4.7 are satisfied. Moreover, either* (a) *or* (b) *holds:*
(a) *h is continuous and $h^\nu(w^\nu) \to h(w) > -\infty$ whenever $w^\nu \to w \in \mathbb{R}^m$. $\forall z \in A$, there exists $z^n \in A \cap Z^n \to z$.*
(b) *$\liminf_\nu h^\nu(w^\nu) \geq h(w) > -\infty$ whenever $w^\nu \to w \in \mathbb{R}^m$.*
 $h^\nu(w) \leq h(w)$ for all $w \in \mathbb{R}^m$, $\nu \in \mathbb{N}$.
 h is lsc and $h(w) \leq h(w')$ for all $w \leq w'$.
 Assumption 4.10 is satisfied for all $n \geq \bar{n}$.
With probability one, the following holds: for any $\varepsilon \in (0, \infty)$, there exists $n_\varepsilon \geq \bar{n}$ such that if $n \geq n_\varepsilon$, $\delta^\nu \to \delta \in [0, \infty)$, and $\{\bar{z}_n^\nu \in \delta^\nu\text{-}\operatorname{argmin} \varphi_n^\nu, \nu \in \mathbb{N}\}$ has a cluster point $\bar{z}_n$, then

$$\varphi\big(T_n(\bar{z}_n)\big) \leq \inf \varphi + \varepsilon + \delta.$$

That is, the optimality gap for $\bar{z}_n$ is $\varepsilon + \delta$.

Proof. Let $n \geq \bar{n}$. Since A is closed and T_n is continuous by Assumption 4.1, it follows that A_n is closed. Likewise, the continuity of f relative to A implies that f_n is continuous relative to A_n.

By Lemmas 4.3 and 4.6, with probability one, any $z_n^\nu \in A_n \to z_n$ as $\nu \to \infty$ implies (4.7). Let $\{\xi_1, \xi_2, \dots\}$ be an event for which this occurs. Since we consider a countable number of values of n, the probability-one set can be assumed to be independent of n.

By the argument in the proof of Theorem 4.9, we obtain that

$$\liminf_\nu \varphi_n^\nu(z_n^\nu) \geq \varphi_n(z_n) \text{ whenever } z_n^\nu \to z_n \text{ as } \nu \to \infty.$$

We next show that for each $\hat{z}_n \in A_n$, there exists $\hat{z}_n^\nu \in A_n \to \hat{z}_n$ as $\nu \to \infty$ such that

$$\limsup_\nu \varphi_n^\nu(\hat{z}_n^\nu) \leq \varphi_n(\hat{z}_n).$$

This holds trivially if $\hat{z}_n \notin A_n$. For $\hat{z}_n \in A_n$, we argue as follows. As in the proof of Theorem 4.9, $z_n^\nu \in A_n \to z_n$ as $\nu \to \infty$ implies that

$$(4.9) \qquad \frac{1}{\nu} \sum_{j=1}^{\nu} G_n^\nu(\xi_j, z_n^\nu) \to \mathbb{E}\big[G_n(\boldsymbol{\xi}, z_n)\big], \quad \text{as } \nu \to \infty.$$

Now, if assumption (a) holds, then set $\hat{z}_n^\nu = \hat{z}_n$ for all ν, which implies that

$$\limsup_\nu \varphi_n^\nu(\hat{z}_n^\nu) = f_n(\hat{z}_n) + \limsup_\nu h^\nu \left(\frac{1}{\nu} \sum_{j=1}^{\nu} G_n^\nu(\xi_j, \hat{z}_n) \right)$$
$$= f_n(\hat{z}_n) + h\Big(\mathbb{E}\big[G_n(\boldsymbol{\xi}, \hat{z}_n)\big] \Big) = \varphi_n(\hat{z}_n).$$

Alternatively, if assumption (b) holds, then there are two cases. (i) If $h(\mathbb{E}[G_n(\boldsymbol{\xi}, \hat{z}_n)]) = \infty$, then we again set $\hat{z}_n^\nu = \hat{z}_n$ for all ν and obtain

$$\limsup_\nu \varphi_n^\nu(\hat{z}_n) \leq \varphi_n(\hat{z}_n)$$

because the right-hand side equals infinity. (ii) If $h(\mathbb{E}[G_n(\boldsymbol{\xi}, \hat{z}_n)]) < \infty$, then, by the first part of Assumption 4.10, there exist $\gamma^k > 0$ and $\bar{z}_n^k \in A_n \to \hat{z}_n$ as $k \to \infty$ such that

$$\left(\mathbb{E}\big[G_n(\boldsymbol{\xi}, \bar{z}_n^k)\big]\right)_i + \gamma^k \le \left(\mathbb{E}\big[G_n(\boldsymbol{\xi}, \hat{z}_n)\big]\right)_i, \ \ i = 1, \dots, m, \ k \in \mathbb{N}.$$

For each k, there exists by (4.9) ν_k such that for all $\nu \ge \nu_k$ one has

$$\left(\frac{1}{\nu} \sum_{j=1}^{\nu} G_n^{\nu}(\xi_j, \bar{z}_n^k)\right)_i \le \left(\mathbb{E}\big[G_n(\boldsymbol{\xi}, \bar{z}_n^k)\big]\right)_i + \gamma^k, \ \ i = 1, \dots, m.$$

These inequalities mean that, for each ν, there is $k(\nu)$ such that

$$\left(\frac{1}{\nu} \sum_{j=1}^{\nu} G_n^{\nu}(\xi_j, \bar{z}_n^{k(\nu)})\right)_i \le \left(\mathbb{E}\big[G_n(\boldsymbol{\xi}, \hat{z}_n)\big]\right)_i, \ \ i = 1, \dots, m,$$

and this can be done such that $k(\nu) \to \infty$ as $\nu \to \infty$. Set $\hat{z}_n^{\nu} = \bar{z}_n^{k(\nu)}$, which then tends to $\hat{z}_n$ as $\nu \to \infty$. These facts and the assumption on h^{ν} and h imply that

$$\text{limsup}_{\nu} \, h^{\nu}\left(\frac{1}{\nu} \sum_{j=1}^{\nu} G_n^{\nu}(\xi_j, \hat{z}_n^{\nu})\right) \le \text{limsup}_{\nu} \, h\left(\frac{1}{\nu} \sum_{j=1}^{\nu} G_n^{\nu}(\xi_j, \hat{z}_n^{\nu})\right) \le h\left(\mathbb{E}[G_n(\boldsymbol{\xi}, \hat{z}_n)]\right)$$

and then also

$$\text{limsup}_{\nu} \, \varphi_n^{\nu}(\hat{z}_n^{\nu}) \le \text{limsup}_{\nu} \, f_n(\hat{z}_n^{\nu}) + \text{limsup}_{\nu} \, h^{\nu}\left(\frac{1}{\nu} \sum_{j=1}^{\nu} G_n^{\nu}(\xi_j, \hat{z}_n^{\nu})\right)$$

$$\le f_n(\hat{z}_n) + h\left(\mathbb{E}\big[G_n(\boldsymbol{\xi}, \hat{z}_n)\big]\right) = \varphi_n(\hat{z}_n).$$

Consequently, with probability one, $\varphi_n^{\nu} \xrightarrow{e} \varphi_n$ as $\nu \to \infty$; see Definition 3.1.

Next, we confirm that $\varphi + \iota_{Z^n} \xrightarrow{e} \varphi$ as $n \to \infty$. Let $z^n \to z$. If $z \notin A$, then $z^n \notin A$ for sufficiently large n because A is closed. Thus,

(4.10) $$\qquad\qquad \text{liminf}_n \left(\varphi(z^n) + \iota_{Z^n}(z^n)\right) \ge \varphi(z)$$

due to the fact that both sides equal infinity. If $z \in A$, then we argue as follows. If z^n remains outside A and/or Z^n, then the same inequality trivially holds because $\varphi(z^n) + \iota_{Z^n}(z^n) = \infty$. Thus, we assume without loss of generality that $z^n \in A \cap Z^n$. Now,

$$\text{liminf}_n \left(\varphi(z^n) + \iota_{Z^n}(z^n)\right) = \text{liminf}_n \left(f(z^n) + h\left(\mathbb{E}\big[G(\boldsymbol{\xi}, z^n)\big]\right)\right)$$

$$\ge f(z) + \text{liminf}_n \, h\left(\mathbb{E}\big[G(\boldsymbol{\xi}, z^n)\big]\right).$$

In view of Lemma 4.8, $\mathbb{E}[G(\boldsymbol{\xi}, z^n)] \to \mathbb{E}[G(\boldsymbol{\xi}, z)]$. Under assumption (a), h is continuous so that the liminf expression tends to $h(\mathbb{E}[G(\boldsymbol{\xi}, z)])$. Under assumption (b), h is lsc and the same expression is bounded from below by $h(\mathbb{E}[G(\boldsymbol{\xi}, z)])$. In any case, we have confirmed that (4.10) holds.

Let $z \in Z$. We also need to construct $z^n \to z$ such that

(4.11) $$\qquad\qquad \text{limsup}_n \left(\varphi(z^n) + \iota_{Z^n}(z^n)\right) \le \varphi(z).$$

We can assume that $z \in A$ because otherwise the right-hand side will be infinity. Under assumption (a), there is $z^n \in A \cap Z^n \to z$, which then implies that

$$\limsup_n \big(\varphi(z^n) + \iota_{Z^n}(z^n)\big) = \limsup_n \Big(f(z^n) + h\big(\mathbb{E}[G(\boldsymbol{\xi}, z^n)]\big)\Big)$$
$$\leq f(z) + \limsup_n h\Big(\mathbb{E}[G(\boldsymbol{\xi}, z^n)]\Big) = \varphi(z),$$

where we again leverage Lemma 4.8. Under assumption (b), we consider two cases. If $h(\mathbb{E}[G(\boldsymbol{\xi}, z)]) = \infty$, then $\varphi(z) = \infty$ and (4.11) holds trivially. If $h(\mathbb{E}[G(\boldsymbol{\xi}, z)]) < \infty$, then, by the second part of Assumption 4.10, there is $z^n \in A \cap Z^n \to z$ such that $\mathbb{E}[G(\boldsymbol{\xi}, z^n)] \leq \mathbb{E}[G(\boldsymbol{\xi}, z)]$. Since $h(w) \leq h(w')$ when $w \leq w'$, this implies that

$$\limsup_n \big(\varphi(z^n) + \iota_{Z^n}(z^n)\big) \leq f(z) + \limsup_n h\Big(\mathbb{E}[G(\boldsymbol{\xi}, z^n)]\Big)$$
$$\leq f(z) + h\Big(\mathbb{E}[G(\boldsymbol{\xi}, z)]\Big) = \varphi(z).$$

We can conclude that $\varphi + \iota_{Z^n} \xrightarrow{e} \varphi$; cf. Definition 3.1.

We have satisfied the requirements of Theorem 3.2 and the conclusion follows. □

Theorem 4.11 supplements Theorem 4.9 by imposing conditions under which the Approximation Algorithm for PDE-Constrained OUU produces a solution with a specific optimality gap. The optimality gap consists of ε, which can be made arbitrarily small at the expense of larger approximating problems in Step 1, and δ, the tolerance invoked in Step 1.

The main additional assumptions relate to h and its approximation h^ν as well as Z^n. Naturally, Z^n needs to approximate Z arbitrarily well as $n \to \infty$ and this suffices if h is continuous; see (a) in the theorem. Trivially, h is continuous when $h(w) = w \in \mathbb{R}$ as in Example 2.1 and when h models finite weights and penalties applied to the various expectations. Continuity fails, however, when the monitoring function models inequality constraints by means of indicator functions. These situations are addressed via (b), where Z^n in interplay with the expectation functions allows one to approach a feasible point along points that are *strictly* feasible; see Assumption 4.10. This is a constraint qualification that resembles the Slater condition from convex optimization. In particular, (b) is tailored to situations when $h(w) = \iota_{(-\infty,0]^m}(w)$ with the penalization approximation $h^\nu(w) = \theta^\nu \sum_{i=1}^m (\max\{0, w_i\})^2$, where $\theta^\nu \to \infty$.

The main practical take away from the theorem is that one should not shy away from "complicated" formulations involving multiple expectation functions appearing in the objective function and in the constraints and even as part of compositions with other functions. Under relatively mild assumptions, the approximating optimization problems arising from discretization, sampling, penalization, smoothing, and other inaccuracies indeed produce solutions with optimality gap no more than ε, which can be arbitrarily small, plus the optimization error δ.

4.3. Buffered failure probability constraint. We return to Example 2.4 and the problem (2.3) with a buffered failure probability constraint. As discussed in subsection 2.2, we aim to solve (2.3) via the reformulation (2.5) and the approximation (2.7). Since the setting is more specific and also slightly different, we refine and adjust Theorem 4.11.

The problem (2.5) is defined using a space $\hat{Z}$, which is assumed to be a separable Banach space, but is augmented with $\mathbb{R}^2$ to produce $Z = \hat{Z} \times \mathbb{R}^2$. Using the product norm, this defines $(Z, \|\cdot\|_Z)$. While the underlying PDE only depends on $\hat{z} \in \hat{Z}$, we retain the notation $s : \Xi \times Z \to U$, which then involves a trivial extension. Likewise,

the functions $\hat{g}_1$ and $\hat{g}_2$ from Example 2.4 are extended to $U \times Z$. Thus, we are in the setting of section 4, with $m = 2$, f given by $\hat{f}$, and $h(w) = w_1 + \iota_{\{0\}}(w_2)$. The approximating problems have $h^\nu(w) = w_1 + y^\nu w_2 + \theta^\nu w_2^2$ for parameters $y^\nu \in \mathbb{R}$ and $\theta^\nu \in (0, \infty)$, and

$$g_1^\nu(u, z) = g_1(u, z),$$
$$g_2^\nu(u, z) = \sigma + \gamma + \frac{1}{1-\alpha} \operatorname{smax}\left(\hat{g}_2(u, z) - \gamma; \beta^\nu\right), \quad \text{with } z = (\hat{z}, \gamma, \sigma).$$

The constraint qualification in Assumption 4.10 can now be simplified.

Assumption 4.12 (qualification for buffered constraint). With the notation

$$\psi(z) = \mathbb{E}\left[\sigma + \gamma + \frac{1}{1-\alpha} \max\left\{0, \hat{g}_2\big(s(\boldsymbol{\xi}, z), z\big) - \gamma\right\}\right], \quad z = (\hat{z}, \gamma, \sigma),$$

suppose that the following hold:

For fixed n and all $z_n \in A_n$, with $\psi(T_n(z_n)) = 0$, there exists $z_n^\nu \in A_n \to z_n$ as $\nu \to \infty$ such that $\psi(T_n(z_n^\nu)) < 0$ for all $\nu \in \mathbb{N}$.

For all $z \in A$, with $\psi(z) = 0$, there exists $z^n \in A \cap Z^n \to z$ as $n \to \infty$ such that $\psi(z^n) < 0$ for $n \in \mathbb{N}$.

PROPOSITION 4.13 (optimality gap under buffered probability constraint). *For fixed $\bar{n} \in \mathbb{N}$, suppose that (2.5) has a finite minimum value τ, A is closed, f is continuous relative to A, the sample $\{\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \dots\}$ is i.i.d. as $\boldsymbol{\xi}$, and, for all $n \geq \bar{n}$, $\inf \varphi_n < \infty$, and suppose that Assumptions 4.1, 4.2, 4.5, 4.7, and 4.12 are satisfied. Moreover, let $\{y^\nu, \nu \in \mathbb{N}\}$ be bounded, $\theta^\nu \to \infty$, $\beta^\nu \to 0$, and φ_n^ν be the objective function in (2.7).*

Then, with probability one, the following hold: for any $\varepsilon \in (0, \infty)$, there exists $n_\varepsilon \geq \bar{n}$ such that if $n \geq n_\varepsilon$, $\delta^\nu \to \delta \in [0, \infty)$, and $\{\bar{z}_n^\nu \in \delta^\nu\text{-} \operatorname{argmin} \varphi_n^\nu, \nu \in \mathbb{N}\}$ has a cluster point $\bar{z}_n$, then $\bar{z} = T_n(\bar{z}_n) \in A$ and

$$f(\bar{z}) + \mathbb{E}\left[\hat{g}_1\big(s(\boldsymbol{\xi}, \bar{z}), \bar{z}\big)\right] \leq \tau + \varepsilon + \delta, \quad \text{b-prob}\left(\hat{g}_2\big(s(\boldsymbol{\xi}, \bar{z}), \bar{z}\big) > 0\right) \leq 1 - \alpha.$$

That is, $\bar{z}$ is feasible in (2.3) with optimality gap of $\varepsilon + \delta$.

Proof. We note that Assumption 4.4 holds automatically with $\varepsilon^\nu = 2\beta^\nu/(1-\alpha)$; see the discussion in subsection 2.2.

Fix $n \geq \bar{n}$. Following the arguments in the proofs of Theorems 4.9 and 4.11, we obtain that

$$\liminf_\nu \varphi_n^\nu(z_n^\nu) \geq \varphi_n(z_n) \quad \text{whenever } z_n^\nu \to z_n \text{ as } \nu \to \infty$$

provided that $\liminf_\nu h^\nu(w^\nu) \geq h(w) > -\infty$ for $w^\nu \to w = (w_1, w_2)$. If $w_2 \neq 0$, then $h(w) = \infty$ and $h^\nu(w^\nu) \to \infty$ because $\{y^\nu, \nu \in \mathbb{N}\}$ is bounded and $\theta^\nu \to \infty$. If $w_2 = 0$, then $h(w) = w_1$ and $h^\nu(w^\nu) \geq w_1^\nu + y^\nu w_2^\nu \to w_1$ after again using the boundedness of y^ν. Thus, the liminf-requirement holds.

We next show that for each $\breve{z}_n \in A_n$, there exists $\breve{z}_n^\nu \in A_n \to \breve{z}_n$ such that

$$(4.12) \qquad\qquad\qquad \limsup_\nu \varphi_n^\nu(\breve{z}_n^\nu) \leq \varphi_n(\breve{z}_n).$$

This holds trivially if $\breve{z}_n \notin A_n$. For $\breve{z}_n = (\hat{z}_n, \gamma, \sigma) \in A_n$, we argue as follows. Let $\{\xi_1, \xi_2, \dots\}$ be an event for which the convergence in (4.7) holds. As in the proof of Theorem 4.9, we know that

$$(4.13) \qquad \frac{1}{\nu}\sum_{j=1}^{\nu} G_n^\nu(\xi_j, z_n^\nu) \to \mathbb{E}\big[G_n(\boldsymbol{\xi}, z_n)\big] \quad \text{whenever } z_n^\nu \in A_n \to z_n.$$

We consider two cases. Let ψ be as defined in Assumption 4.12.

First, suppose that $\psi(T_n(\check{z}_n)) \neq 0$. Then, we set $\check{z}_n^\nu = \check{z}_n$ for all ν and obtain (4.12) because the right-hand side equals infinity.

Second, suppose that $\psi(T_n(\check{z}_n)) = 0$. By Assumption 4.12, there exist $z_n^k = (\bar{z}_n^k, \gamma^k, \sigma^k) \in A_n \to \check{z}_n$ as $k \to \infty$ such that $\psi(T_n(z_n^k)) < 0$. Thus, for each k, there exists by (4.13) ν_k such that for all $\nu \geq \nu_k$ one has

$$\sigma^k + \gamma^k + \frac{1}{1-\alpha}\frac{1}{\nu}\sum_{j=1}^\nu \mathrm{smax}\left(\hat{g}_2\Big(s\big(\xi_j, T_n(z_n^k)\big), T_n(z_n^k)\Big) - \gamma^k;\, \beta^\nu\right) \leq \frac{1}{2}\psi\big(T_n(z_n^k)\big).$$

This inequality means that, for each ν, there is $k(\nu)$ such that

$$\sigma^{k(\nu)} + \gamma^{k(\nu)} + \frac{1}{1-\alpha}\frac{1}{\nu}\sum_{j=1}^\nu \mathrm{smax}\left(\hat{g}_2\Big(s\big(\xi_j, T_n(z_n^{k(\nu)})\big), T_n(z_n^{k(\nu)})\Big) - \gamma^{k(\nu)};\, \beta^\nu\right) \leq 0$$

and this can be done such that $k(\nu) \to \infty$ as $\nu \to \infty$. Set $\tilde{\sigma}^\nu \geq 0$ equal to the negative of the left-hand side in this inequality. Thus, $\check{z}_n^\nu = (\bar{z}_n^{k(\nu)}, \gamma^{k(\nu)}, \sigma^{k(\nu)} + \tilde{\sigma}^\nu) \in A_n$ and tends to $\check{z}_n$ as $\nu \to \infty$. Moreover,

$$h^\nu\left(\frac{1}{\nu}\sum_{j=1}^\nu G_n^\nu(\xi_j, \check{z}_n^\nu)\right) = \frac{1}{\nu}\sum_{j=1}^\nu \hat{g}_1\Big(s\big(\xi_j, T_n(\check{z}_n^\nu)\big), T_n(\check{z}_n^\nu)\Big) + y^\nu \cdot 0 + \theta^\nu \cdot 0^2$$
$$\to \mathbb{E}\Big[\hat{g}_1\big(s(\boldsymbol{\xi}, T_n(\check{z}_n)), T_n(\check{z}_n)\big)\Big] = h\Big(\mathbb{E}\big[G_n(\boldsymbol{\xi}, \check{z}_n)\big]\Big).$$

Since $\check{z}_n^\nu \in A_n$ and f_n is continuous on A_n, one obtains $\varphi_n^\nu(\check{z}_n^\nu) \to \varphi_n(\check{z}_n)$. Consequently, with probability one, we have established that $\varphi_n^\nu \xrightarrow{e} \varphi_n$ as $\nu \to \infty$; see Definition 3.1.

Next, we confirm that $\varphi + \iota_{Z^n} \xrightarrow{e} \varphi$ as $n \to \infty$. Let $z^n \to z$. Since h is lsc, we can argue as in the proof of Theorem 4.11 and conclude that

$$\mathrm{liminf}_n\big(\varphi(z^n) + \iota_{Z^n}(z^n)\big) \geq \varphi(z).$$

Let $z = (\hat{z}, \gamma, \sigma) \in Z$. We also need to construct $z^n \to z$ such that

(4.14) $$\mathrm{limsup}_n\big(\varphi(z^n) + \iota_{Z^n}(z^n)\big) \leq \varphi(z).$$

We can assume that $z \in A$ because otherwise the right-hand side will be infinity. If $\psi(z) \neq 0$, then $\varphi(z) = \infty$ and (4.14) holds trivially again. If $\psi(z) = 0$, then, by the second part of Assumption 4.12, there is $(\hat{z}^n, \gamma^n, \sigma^n) \in A \cap Z^n \to z$ such that $\psi((\hat{z}^n, \gamma^n, \sigma^n)) < 0$. Set $\tilde{\sigma}^n = -\psi((\hat{z}^n, \gamma^n, \sigma^n))$, which then vanishes as $n \to \infty$, and also construct $z^n = (\hat{z}^n, \gamma^n, \sigma^n + \tilde{\sigma}^n)$. Since $\psi(z^n) = 0$, we conclude that

$$\varphi(z^n) + \iota_{Z^n}(z^n) = f(z^n) + \mathbb{E}\Big[\hat{g}_1\big(s(\boldsymbol{\xi}, z^n), z^n\big)\Big] \to f(z) + \mathbb{E}\Big[\hat{g}_1\big(s(\boldsymbol{\xi}, z), z\big)\Big] = \varphi(z)$$

after invoking Lemma 4.8. We have established that $\varphi + \iota_{Z^n} \xrightarrow{e} \varphi$; cf. Definition 3.1. The assumptions of Theorem 3.2 therefore holds and $\varphi(T_n(\bar{z}_n)) \leq \inf \varphi + \varepsilon + \delta$. Since $\inf \varphi = \tau$ is finite, $T_n(\bar{z}_n)$ is feasible in (2.5). The conclusion then follows by reversing the arguments leading from (2.3) to (2.4) and to (2.5). □

The proposition shows that a cluster point constructed by the approximating problems (2.7) produces a feasible control for the actual problem that is suboptimal with tolerance $\varepsilon + \delta$. There is much flexibility in choosing the multipliers y^ν. A penalty method remains possible by setting $y^\nu = 0$, but update rules for y^ν from augmented Lagrangian methods may be computationally superior. Under additional assumptions, we conjecture that "exactness" is possible and then θ^ν can remain bounded; we omit the details and refer the reader to [38, section 6.B].

5. Application. This section examines the assumptions of section 4 in the context of a distributed optimal control problems. Specifically, for $d = 1, 2$, or 3, let $D \subset \mathbb{R}^d$ be a bounded open set representing a physical domain, and let ∂D be the corresponding boundary. Set $\bar{D} = D \cup \partial D$. We denote by $L^2(D)$ and $L^2(\partial D)$ the spaces of Lebesgue square-integrable functions from D and ∂D to $\mathbb{R}$, respectively. As before, $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space. We consider thermal conduction modeled by the following PDE with random conductivity coefficient: For a.e. $\omega \in \Omega$,

$$(5.1) \qquad \begin{aligned} -\nabla \cdot (\boldsymbol{\xi}(\omega)\nabla u) &= c_1 z \quad \text{in } D, \\ \boldsymbol{\xi}(\omega)\nabla u \cdot \vec{n} &= c_2(s_e - u) \quad \text{on } \partial D, \end{aligned}$$

where the state variable $u : \bar{D} \to \mathbb{R}$ represents the unknown temperature field, $s_e \in L^2(\partial D)$ is a given exterior temperature along the boundary, $z \in Z = L^2(D)$ is the control in a nonempty admissible set $A = \{z \in Z \mid \underline{z} \le z(x) \le \bar{z} \text{ for a.e. } x \in D\}$, $c_1 : D \to [0, \infty)$ and $c_2 : \partial D \to [0, \infty)$ are known coefficients that are uniformly bounded from above, $\vec{n}$ is a unit normal vector pointing outside $\bar{D}$, and $\boldsymbol{\xi}(\omega) : \bar{D} \to (0, \infty)$ is a thermal conductivity field. Specifically,

$$\boldsymbol{\xi}(\omega)(x) = \exp\left(b_0(x) + \sum_{j=1}^{J} b_j(x)\mathbf{y}_j(\omega)\right),$$

where $J \in \mathbb{N}$, $\mathbf{y}_1, \ldots, \mathbf{y}_J$ are i.i.d. random variables defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $b_0, \ldots, b_J : \bar{D} \to \mathbb{R}$ are given functions. This induces a probability space $(\Xi, \mathcal{B}, P)$ as discussed in section 2, where elements $\xi \in \Xi$ are functions from $\bar{D}$ to $(0, \infty)$. One example of such a random field has $\log(\boldsymbol{\xi})$ given by a truncated Kahunen–Loève expansion of a Gaussian random field $\mathcal{N}(b_0, \mathcal{C})$ with mean b_0 and covariance $\mathcal{C}$ [39], where $b_j = \sqrt{\lambda_j}\varphi_j$, (λ_j, φ_j) is an eigenpair of $\mathcal{C}$, and $\mathbf{y}_1, \ldots, \mathbf{y}_J$ follow the standard normal distribution. In another example, $\boldsymbol{\xi}$ represents a piecewise random thermal conductivity coefficient with b_j denoting a rescaled characteristic function supported on a subdomain $D_j \subset \bar{D}$.

Assumption 5.1. There exist $\underline{c}, \bar{c} : \Xi \to (0, \infty)$ such that

$$(5.2) \qquad 0 < \underline{c}(\xi) = \operatorname*{ess\,inf}_{x \in \bar{D}} \xi(x) \le \operatorname*{ess\,sup}_{x \in \bar{D}} \xi(x) = \bar{c}(\xi) < \infty \quad \text{for a.e. } \xi \in \Xi,$$

and, regardless of $0 \le p \le 1$ and $0 \le q \le 3$, one has

$$(5.3) \qquad \int_\Xi \frac{(\bar{c}(\xi))^p}{(\underline{c}(\xi))^q} dP(\xi) < \infty.$$

Following the notation and terminology from (2.1), we consider $U = L^2(D)$ and two performance functions: the discrepancy between the solution and a desired temperature field $s_d : D \to \mathbb{R}$ and the shortfall of the solution average over $D_t \subset D$ relative to a threshold temperature s_t. They are captured by $g_1, g_2 : U \times Z \to \mathbb{R}$ with

$$(5.4) \qquad g_1(u, z) = \int_D (u(x) - s_d(x))^2 dx, \quad g_2(u, z) = s_t - \int_{D_t} u(x)dx.$$

We adopt the cost function $f(z) = \theta\|z\|_Z^2$ for some $\theta > 0$, which is continuous, and thus satisfying the continuity assumption of f in Theorem 4.11 and Proposition 4.13.

Let $V = H^1(D) = \{v \in L^2(D) \mid |\nabla v| \in L^2(D)\}$ with the norm $\|v\|_V = \|v\|_{L^2(D)} + \||\nabla v|\|_{L^2(D)}$. For fixed $\xi \in \Xi$ and $z \in Z$, $u \in V$ is a weak solution of (5.1) if

$$(5.5) \qquad a(u, v; \xi) = \ell(v) + b(z, v) \quad \forall v \in V,$$

where

$$a(u, v; \xi) = \int_D \xi(x) \nabla u(x) \cdot \nabla v(x) dx + \int_{\partial D} c_2(x) u(x) v(x) dx \quad \forall u, v \in V,$$

$$b(z, v) = \int_D c_1(x) z(x) v(x) dx, \text{ and } \ell(v) = \int_{\partial D} c_2(x) s_e(x) v(x) dx \quad \forall v \in V.$$

A solution u of (5.5) defines a mapping from $\Xi \times Z$ to V. Since the performance functions g_1, g_2 are defined on $L^2(D) \times L^2(D)$, it becomes more natural to adopt $U = L^2(D)$ as the range space for the solution mapping. Thus, the solution mapping $s : \Xi \times Z \to U$ is given as $s(\xi, z) = u$, where u is the solution of (5.5) under input $\xi \in \Xi$ and $z \in Z$. We are in the setting of the earlier sections with $Z = U = L^2(D)$.

5.1. Approximations. We use a finite element method to approximate solutions and controls. Let X_h^p denote a finite element space of piecewise polynomials of total degree $p \in \mathbb{N} \cup \{0\}$ on each element K of a triangulation $\mathcal{T}_h$ with mesh size h. For simplicity, we set $Z^n = X_h^0$ for some h such that n represents the number of elements. Let $T_n : \mathbb{R}^n \to Z^n$ be given by

$$(5.6) \qquad z^n = T_n(z_n) = \sum_{i=1}^n z_n^i \psi_i(\cdot) \quad \text{for } z_n = (z_n^1, z_n^2, \ldots, z_n^n) \in \mathbb{R}^n,$$

where $\psi_i(x) = 1$ for x in element K_i and zero otherwise. Let $V^\nu = X_h^1$ for some h depending on an index ν, with r^ν representing the number of finite element basis functions in V^ν, i.e., the number of degrees of freedom. Although not common in practice, Z^n and V^ν can be constructed with different meshes. The Galerkin approximation of the problem (5.5) reads: for $\xi \in \Xi$ and control $z^n \in Z^n$, find $u^\nu \in V^\nu$ such that

$$(5.7) \qquad a(u^\nu, v^\nu; \xi) = \ell(v^\nu) + b(z^n, v^\nu) \quad \forall v^\nu \in V^\nu.$$

We remark that (5.7) can be extended to any control $z \in Z$. The approximating solution mapping $s^\nu : \Xi \times Z \to U$ is therefore defined as $s^\nu(\xi, z) = u^\nu$, where u^ν is the solution of (5.7) under input $\xi \in \Xi$ and $z \in Z$.

These approximations have the following properties.

PROPOSITION 5.2. *The mapping $T_n : \mathbb{R}^n \to Z^n$ in (5.6) is continuous, with*

$$(5.8) \qquad \left\| T_n(z_n) \right\|_Z \leq c_K^{1/2} h^{d/2} \|z_n\|_2 \quad \forall z_n \in \mathbb{R}^n,$$

for a constant $c_K < \infty$ independent of h, n, and z_n. Moreover, for any $z \in A$, there exists $\{z_n \in \mathbb{R}^n, n \in \mathbb{N}\}$ such that

$$(5.9) \qquad \left\| z - T_n(z_n) \right\|_Z \to 0 \text{ as } n \to \infty.$$

Proof. By definition of the piecewise constant finite element approximation, we have

$$
\begin{aligned}
\left\| T_n(z_n) \right\|_Z^2 &= \int_D \left(T_n(z_n)(x) \right)^2 dx = \sum_{i=1}^n \sum_{j=1}^n z_n^i z_n^j \int_D \psi_i(x) \psi_j(x) dx \\
&= \sum_{i=1}^n (z_n^i)^2 |K_i| \leq c_K h^d \|z_n\|_2^2,
\end{aligned}
$$

$$(5.10)$$

where the inequality holds because each element $K_i \in \mathcal{T}_h$ of mesh size h has length, area, or volume $|K_i| \leq c_K h^d$ in $d = 1, 2, 3$, respectively, for some constant $c_K < \infty$ independent of n. Moreover, for any function $z \in A$, which is measurable, bounded, and supported in the bounded domain D, the simple function approximation $T_n(z_n)$ is uniformly convergent. □

PROPOSITION 5.3. *Suppose that Assumption* 5.1 *holds. For $\xi \in \Xi$ and $z \in Z$, there exist a unique solution $s(\xi, z) \in V$ of* (5.5) *and a unique solution $s^\nu(\xi, z) \in V^\nu \subset V$ of* (5.7). *Moreover, we have the stability estimate*

$$(5.11) \qquad \left\|s(\xi, z)\right\|_V, \left\|s^\nu(\xi, z)\right\|_V \leq c_2^s(\xi) + c_1^s(\xi)\|z\|_Z,$$

where c_2^s and c_1^s are given in (5.19). *Furthermore, there holds the continuity estimate*

$$(5.12) \qquad \left\|s(\xi, z) - s(\xi, z')\right\|_V, \left\|s^\nu(\xi, z) - s^\nu(\xi, z')\right\|_V \leq c_z^s(\xi)\|z - z'\|_Z,$$

where $c_z^s(\xi)$ is given in (5.21), *and the continuity estimate*

$$(5.13) \qquad \left\|s(\xi, z) - s(\xi', z)\right\|_V, \left\|s^\nu(\xi, z) - s^\nu(\xi', z)\right\|_V \leq c_\xi^s(\xi, z)\|\xi - \xi'\|_{L^\infty(D)},$$

where $c_\xi^s(\xi, z)$ is given in (5.23). *Since $V = H^1(D) \subset U = L^2(D)$ and $\|u\|_U \leq \|u\|_V$ for any $u \in V$, all the estimates above also hold in $\|\cdot\|_U$.*

Finally, we have the error estimate

$$(5.14) \qquad \left\|s(\xi, z) - s^\nu(\xi, z)\right\|_U \leq c_\nu^s(\xi)\left\|s(\xi, z)\right\|_V (r^\nu)^{-1/d} \to 0 \text{ as } \nu \to \infty,$$

where $c_\nu^s(\xi)$ is given in (5.25), *r^ν is the number of degrees of freedom of V^ν, which satisfies $r^\nu \to \infty$ as $\nu \to \infty$, and d is the dimension of the physical domain $D \subset \mathbb{R}^d$.*

Proof. The proof follows that for finite element Galerkin approximation of deterministic elliptic PDEs (see, for example, [30, sect. 4]) but is enriched with additional details to facilitate verification of Assumptions 4.2 and 4.5.

By (5.2) and the Poincaré inequality $\|u\|_{L^2(D)} \leq c_P\|\nabla u\|_{L^2(D)}$, with the Poincaré constant $c_P > 0$, one obtains

$$(5.15) \qquad a(u, u; \xi) \geq \underline{c}(\xi)\|\nabla u\|_{L^2(D)}^2 \geq \frac{\underline{c}(\xi)}{\sqrt{1 + c_P}}\|u\|_V^2 \quad \forall u \in V.$$

The Cauchy–Schwarz inequality, (5.2), $\overline{c}_2 = \|c_2\|_{L^\infty(D)}$, and the trace inequality $\|v\|_{L^2(\partial D)} \leq c_T\|v\|_V$ with constant $c_T > 0$ produce

$$(5.16) \qquad a(u, v; \xi) \leq \left(\overline{c}(\xi) + \overline{c}_2 c_T^2\right)\|u\|_V\|v\|_V.$$

Similarly, with $\overline{c}_1 = \|c_1\|_{L^\infty(D)}$ and $\|v\|_{L^2(D)} \leq \|v\|_V$, we have

$$(5.17) \qquad b(z, v) \leq \overline{c}_1\|z\|_Z\|v\|_V \quad \forall z \in Z, v \in V.$$

Again, by the Cauchy–Schwarz inequality and the trace inequality, we reach

$$(5.18) \qquad \ell(v) \leq \overline{c}_2 c_T\|s_e\|_{L^2(\partial D)}\|v\|_V \quad \forall v \in V.$$

Combining the above four inequalities, we conclude that there exist a unique solution $s(\xi, z) \in V$ of (5.5) by the Lax–Milgram theorem and also a unique solution $s^\nu(\xi, z) \in V^\nu$ of (5.7) since $V^\nu \subset V$, for which the above four inequalities also hold. Moreover,

replacing $v = u$ in (5.5) and $v^\nu = u^\nu$ in (5.7) by the above inequalities, we have the stability estimate (5.11), with c_2^s and c_1^s given by

$$(5.19) \qquad c_2^s(\xi) = \frac{\overline{c}_2 c_T \|s_e\|_{L^2(\partial D)} \sqrt{1 + c_P}}{\underline{c}(\xi)} \quad \text{and} \quad c_1^s(\xi) = \frac{\overline{c}_1 \sqrt{1 + c_P}}{\underline{c}(\xi)}.$$

For the continuity of $s(\xi, \cdot)$, we take $z, z' \in Z$ in (5.5) and denote $s = s(\xi, z)$ and $s' = s(\xi, z')$ as the corresponding solutions, respectively. By subtracting the two equations (5.5) at $z, z' \in Z$, we have

$$(5.20) \qquad a(s - s', v; \xi) = b(z - z', v) \quad \forall v \in V.$$

By $v = s - s'$, we obtain the continuity estimate (5.12) using (5.15) and (5.17), with

$$(5.21) \qquad c_z^s(\xi) = \frac{\overline{c}_1 \sqrt{1 + c_P}}{\underline{c}(\xi)}.$$

By the same argument, the same continuity estimate holds for $s^\nu(\xi, \cdot)$.

Similarly, for the continuity of $s(\cdot, z)$, we take $\xi, \xi' \in \Xi$ in (5.5) and denote $s = s(\xi, z)$ and $s' = s(\xi', z)$ as the corresponding solutions, respectively. By subtracting the two equations (5.5) at $\xi, \xi' \in \Xi$, we have

$$(5.22) \qquad \int_D \xi(x) \nabla\big(s(x) - s'(x)\big) \cdot \nabla v(x) dx = \int_D \big(\xi'(x) - \xi(x)\big) \nabla s'(x) \cdot \nabla v(x) dx.$$

By taking $v = s - s'$, we obtain the continuity estimate (5.13) using (5.15) by noting that $c_2 \geq 0$ and the stability estimate (5.11), with $c_\xi^s(\xi, z)$ given by

$$(5.23) \qquad c_\xi^s(\xi, z) = \frac{\sqrt{1 + c_P}}{\underline{c}(\xi)} \big(c_2^s(\xi) + c_1^s(\xi) \|z\|_Z\big).$$

The same holds for the approximation solution s^ν by the same argument.

The error estimate (5.14) follows the proof of [30, sect. 4, Thm. 4.7], which satisfies

$$(5.24) \qquad \big\|s(\xi, z) - s^\nu(\xi, z)\big\|_U \leq h \hat{c} \frac{\overline{c}(\omega)}{\underline{c}(\xi)} \big\|s(\xi, z)\big\|_V,$$

where h is the finite element mesh size for the discretization of the domain D and $\hat{c} \in (0, \infty)$ is a constant, independent of ξ and z. Note that the mesh size h is related to the number of degrees of freedom ν as $h = O((r^\nu)^{-1/d})$ in $D \subset \mathbb{R}^d$, i.e., $h \leq \breve{c}(r^\nu)^{-1/d}$ for some $\breve{c} \in (0, \infty)$. The mesh size is thus tending to zero. Then, we obtain (5.14) with $c_\nu^s(\xi)$ given by

$$(5.25) \qquad c_\nu^s(\xi) = \hat{c} \breve{c} \overline{c}(\xi) / \underline{c}(\xi),$$

which concludes all the estimates. □

To address Assumption 4.1, we note that $s(\xi, \cdot) : Z \to U$ is continuous due to (5.12) and $s(\cdot, z)$ and $s^\nu(\cdot, z)$ are measurable due to (5.13).

PROPOSITION 5.4. *The performance functions g_1 and g_2 defined in (5.4) satisfy the continuity estimates: for all $u, u' \in U, z, z' \in Z$,*

$$(5.26) \qquad \begin{aligned} \big|g_1(u, z) - g_1(u', z')\big| &\leq \big(\|u\|_U + \|u'\|_U + 2\|s_d\|_U\big) \|u - u'\|_U, \\ \big|g_2(u, z) - g_2(u', z')\big| &\leq |D_t| \|u - u'\|_U, \end{aligned}$$

where $|D_t|$ is the size of the subdomain $D_t \subset D$. This implies Assumption 4.1 and Assumption 4.5 for g_1 and g_2. Moreover, for all $u \in U, z \in Z$, the performance functions are bounded from above by

$$\big|g_1(u,z)\big| \leq \big(\|u\|_U + \|s_d\|_U\big)^2, \quad \big|g_2(u,z)\big| \leq |s_t| + |D_t|\|u\|_U.$$

Proof. By definition of g_1 in (5.4), we have

$$\big|g_1(u,z) - g_1(u',z')\big| = \left|\int_D \big(u(x) - s_d(x)\big)^2 - \big(u'(x) - s_d(x)\big)^2 dx\right|$$
$$\leq \big(\|u\|_U + \|u'\|_U + 2\|s_d\|_U\big)\|u - u'\|_U,$$

which concludes the continuity estimate. The upper bound for g_1 is satisfied by the Cauchy–Schwarz inequality. For g_2, we have

$$\big|g_2(u,z) - g_2(u',z')\big| = \left|\int_{D_t} \big(u(x) - u'(x)\big)dx\right| \leq |D_t|\,\|u - u'\|_U,$$

which concludes the continuity estimate. The upper bound for g_2 is also satisfied by the Cauchy–Schwarz inequality. □

Note that the continuity of g_1 and g_2 implies the measurability of $g_1(s(\xi,z),z)$ and $g_2(s(\xi,z),z)$ w.r.t. ξ, by the measurability of the solution $s(\xi,z)$ w.r.t. ξ from Proposition 5.3 and the Doob–Dynkin lemma [28, Lemma 2.12], which verifies Assumption 4.1 on the measurability of $G(\xi,z) = (g_1(s(\xi,z),z), g_2(s(\xi,z),z))$ w.r.t. ξ.

PROPOSITION 5.5. *Let $n \in \mathbb{N}$ be fixed and consider the setting of this section under Assumption 5.1. For $\bar{z}_n \in A_n$, $\rho > 0$, and $z_n \in \mathbb{B}(\bar{z}_n, \rho) \cap A_n$, let $z^n = T_n(z_n) \in Z^n$. For $i = 1, 2$, there exists an integrable function $\kappa_i : \Xi \to \mathbb{R}$, such that*

$$\Big|g_i\big(s(\xi, z^n), z^n\big)\Big| \leq \kappa_i(\xi),$$

which also hold when z^n is replaced by any $z \in A$ such that $\|z - \bar{z}\|_Z \leq \rho$ for $\bar{z} \in A$. Moreover, there exist integrable $\lambda_i : \Xi \to \mathbb{R}$ and $\Delta_i^\nu(\bar{z}_n) \to 0$ as $\nu \to \infty$, such that

$$\Big|g_i\big(s(\xi, z^n), z^n\big) - g_i\big(s^\nu(\xi, z^n), z^n\big)\Big| \leq \lambda_i(\xi)\Delta_i^\nu(\bar{z}_n),$$

where $\kappa_i, \lambda_i, \Delta_i^\nu$ are as defined in the proof.

Proof. For $z_n \in \mathbb{B}(\bar{z}_n, \rho) \cap A_n$, one has $\|z^n\|_Z = \|T_n(z_n)\|_Z < \infty$ by Proposition 5.2.

First, we prove the upper bound and the error estimate for g_1. By definition and Proposition 5.4, we have

$$g_1\big(s(\xi, z^n), z^n\big) \leq \big(\|s(\xi, z^n)\|_U + \|s_d\|_U\big)^2 \leq \big(c_2^s(\xi) + c_1^s(\xi)\|z^n\|_Z + \|s_d\|_U\big)^2,$$

where we used the stability estimate (5.11). The right-hand side expression furnishes a value for $\kappa_1(\xi)$. The random variable κ_1 is integrable if $(c_2^s(\cdot))^2$, $(c_1^s(\cdot))^2$, $c_2^s(\cdot)c_1^s(\cdot)$, $c_2^s(\cdot)$, and $c_2^s(\cdot)$ are all integrable. By definition of $c_2^s(\xi)$ and $c_2^s(\xi)$ in (5.19), this is satisfied because $1/\underline{c}(\cdot)$ and $1/(\underline{c}(\cdot))^2$ are integrable by Assumption 5.1. Moreover, the integrability of κ_1 holds when $z^n \in Z^n$ is replaced by any $z \in Z$ since $\|z\|_Z < \infty$.

By the continuity estimate for g_1 in Proposition 5.4, and the error and stability estimates in Proposition 5.3 for $s(\xi, z^n)$ and $s^\nu(\xi, z^n)$, we have

$$
\begin{aligned}
&\left| g_1\big(s(\xi, z^n), z^n\big) - g_1\big(s^\nu(\xi, z^n), z^n\big) \right| \\
&\quad \leq c_\nu^s(\xi) \left(\left\| s(\xi, z^n) \right\|_U + \left\| s^\nu(\xi, z^n) \right\|_U + 2\|s_d\|_U \right) \|s\|_V (r^\nu)^{-1/d} \\
&\quad \leq c_\nu^s(\xi) \Big((c_2^s(\xi))^2 + 2c_2^s(\xi)c_1^s(\xi)\|z^n\|_Z + (c_1^s(\xi)\|z^n\|_Z)^2 \\
&\qquad + c_2^s(\xi)\|s_d\|_U + c_1^s(\xi)\|s_d\|_U\|z^n\|_Z \Big) (r^\nu)^{-1/d} \leq \lambda_1(\xi)\Delta_1^\nu(\bar{z}_n),
\end{aligned}
$$

where

$$
\lambda_1(\xi) = c_\nu^s(\xi) \Big((c_2^s(\xi))^2 + 2c_2^s(\xi)c_1^s(\xi) + (c_1^s(\xi))^2 + c_2^s(\xi) + c_1^s(\xi) \Big),
$$
$$
\Delta_1^\nu(\bar{z}_n) = \sup_{z_n \in \mathbb{B}(\bar{z}_n, \rho)} \big(1 + \|T_n(z_n)\|_Z + \|T_n(z_n)\|_Z^2 + \|s_d\|_U + \|s_d\|_U\|T_n(z_n)\|_Z\big)(r^\nu)^{-1/d},
$$

which satisfies $\Delta_1^\nu(\bar{z}_n) \to 0$ as $\nu \to \infty$. By the definition of c_1^s and c_2^s in (5.19), we have that λ_1 is integrable as long as $\bar{c}(\cdot)/(\underline{c}(\cdot))^3$ and $\bar{c}(\cdot)/(\underline{c}(\cdot))^2$ are integrable, which is the case by Assumption 5.1.

Second, for g_2, by Proposition 5.4 we have

$$
\left| g_2\big(s(\xi, z^n), z^n\big) \right| \leq |s_t| + |D_t| \big\| s(\xi, z^n) \big\|_U \leq |s_t| + |D_t| \big(c_2^s(\xi) + c_1^s(\xi)\|z^n\|_Z \big),
$$

where we used the stability estimate (5.11), with $|D_t|$ measuring the size of D_t. The right-hand side furnishes an expression for $\kappa_2(\xi)$. This defines an integrable random variable κ_2 because $c_1^s(\cdot)$ and $c_2^s(\cdot)$ in (5.19) are integrable as a result of $1/\underline{c}(\cdot)$ being integrable by Assumption 5.1. Moreover, the integrability of κ_2 holds when $z^n \in Z^n$ is replaced by $z \in Z$ as $\|z\|_Z < \infty$.

By the continuity estimate for g_2 in Proposition 5.4, and the error and stability estimates in Proposition 5.3 for $s(\xi, z^n)$ and $s^\nu(\xi, z^n)$, we have

$$
\begin{aligned}
\left| g_2\big(s(\xi, z^n), z^n\big) - g_2\big(s^\nu(\xi, z^n), z^n\big) \right| &\leq |D_t|c_\nu^s(\xi)\big(c_2^s(\xi) + c_1^s(\xi)\|z^n\|_Z\big)(r^\nu)^{-1/d} \\
&\leq \lambda_2(\xi)\Delta_2^\nu(\bar{z}_n),
\end{aligned}
$$

where $\lambda_2(\xi) = |D_t|c_\nu^s(\xi)(c_2^s(\xi) + c_1^s(\xi))$, which is integrable as long as $\bar{c}(\cdot)/(\underline{c}(\cdot))^2$ is integrable, which holds by Assumption 5.1. Moreover,

$$
\Delta_2^\nu(\bar{z}_n) = \sup_{z_n \in \mathbb{B}(\bar{z}_n, \rho)} \big(1 + \|T_n(z_n)\|_Z\big)(r^\nu)^{-1/d},
$$

which satisfies $\Delta_2^\nu(\bar{z}_n) \to 0$ as $\nu \to \infty$. $\qquad \square$

To this end, all the assumptions made in section 4.1 are satisfied under Assumption 5.1 for the example presented in this section.

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. Uncertain. Quantif., 5 (2017), pp. 1166–1192, https://doi.org/10.1137/16M106306X.
[2] A. A. ALI, E. ULLMANN, AND M. HINZE, *Multilevel Monte Carlo analysis for optimal control of elliptic PDEs with random coefficients*, SIAM/ASA J. Uncertain. Quantif., 5 (2017), pp. 466–492, https://doi.org/10.1137/16M109870X.

[3] P. BENNER, A. ONWUNTA, AND M. STOLL, *Block-diagonal preconditioning for optimal control problems constrained by PDEs with uncertain inputs*, SIAM J. Matrix Anal. Appl., 37 (2016), pp. 491–518, https://doi.org/10.1137/15M1018502.

[4] A. SUR AND J. R. BIRGE, *Epi-convergence of sample averages of a random lower semi-continuous functional generated by a Markov chain and application to stochastic optimization*, Optim. Online, 2020 (2020).

[5] J.-E. BYUN AND J. O. ROYSET, *Data-driven optimization of reliability using buffered failure probability*, Struct. Saf., 98 (2022), 102232.

[6] P. CHEN AND A. QUARTERONI, *Weighted reduced basis method for stochastic optimal control problems with elliptic PDE constraint*, SIAM/ASA J. Uncertain. Quantif., 2 (2014), pp. 364–396, https://doi.org/10.1137/130940517.

[7] P. CHEN, A. QUARTERONI, AND G. ROZZA, *Multilevel and weighted reduced basis method for stochastic optimal control problems constrained by Stokes equations*, Numer. Math., 133 (2016), pp. 67–102.

[8] P. CHEN AND O. GHATTAS, *Taylor approximation for chance constrained optimization problems governed by partial differential equations with high-dimensional random parameters*, SIAM/ASA J. Uncertain. Quantif., 9 (2021), pp. 1381–1410, https://doi.org/10.1137/20M1381381.

[9] P. CHEN, M. HABERMAN, AND O. GHATTAS, *Optimal design of acoustic metamaterial cloaks under uncertainty*, J. Comput. Phys., 431 (2021), 110114.

[10] P. CHEN, A. QUARTERONI, AND G. ROZZA, *Stochastic optimal Robin boundary control problems of advection-dominated elliptic equations*, SIAM J. Numer. Anal., 51 (2013), pp. 2700–2722, https://doi.org/10.1137/120884158.

[11] P. CHEN, U. VILLA, AND O. GHATTAS, *Taylor approximation and variance reduction for PDE-constrained optimal control under uncertainty*, J. Comput. Phys., 385 (2019), pp. 163–186.

[12] H. T. H. DIEM, *Consistency of statistical estimators of solutions to stochastic optimization problems*, J. Global Optim., 83 (2022), pp. 825–842.

[13] S. GARREIS, T. M. SUROWIEC, AND M. ULBRICH, *An interior-point approach for solving risk-averse PDE-constrained optimization problems with coherent risk measures*, SIAM J. Optim., 31 (2021), pp. 1–29, https://doi.org/10.1137/19M125039X.

[14] C. GEIERSBACH, E. LOAYZA-ROMERO, AND K. WELKER, *Stochastic approximation for optimization in shape spaces*, SIAM J. Optim., 31 (2021), pp. 348–376, https://doi.org/10.1137/20M1316111.

[15] P. A. GUTH, V. KAARNIOJA, F. Y. KUO, C. SCHILLINGS, AND I. H. SLOAN, *A Quasi-Monte Carlo Method for an Optimal Control Problem under Uncertainty*, arXiv:1910.10022, 2019.

[16] W. W. HAGER, *Runge-Kutta methods in optimal control and the transformed adjoint system*, Numer. Math., 87 (2000), pp. 247–282.

[17] M. HINZE, R. PINNAU, M. ULBRICH, AND S. ULBRICH, *Optimization with PDE Constraints*, Math. Model. Theory Appl. 23, Springer, Cham, 2009.

[18] 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.

[19] 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 J. Sci. Comput., 35 (2013), pp. A1847–A1879, https://doi.org/10.1137/120892362.

[20] D. P. KOURI AND T. M. 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.

[21] D. P. KOURI AND T. M. SUROWIEC, *Existence and optimality conditions for risk-averse PDE-constrained optimization*, SIAM/ASA J. Uncertain. Quantif., 6 (2018), pp. 787–815, https://doi.org/10.1137/16M1086613.

[22] D. P. KOURI AND T. M. SUROWIEC, *Epi-regularization of risk measures*, Math. Oper. Res., 45 (2020), pp. 774–795.

[23] A. KUNOTH AND C. SCHWAB, *Sparse adaptive tensor Galerkin approximations of stochastic PDE-constrained control problems*, SIAM/ASA J. Uncertain. Quantif., 4 (2016), pp. 1034–1059, https://doi.org/10.1137/15M1041390.

[24] 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.

[25] A. MAFUSALOV AND S. URYASEV, *Buffered probability of exceedance: Mathematical properties and optimization*, SIAM J. Optim., 28 (2018), pp. 1077–1103, https://doi.org/10.1137/15M1042644.

[26] M. MARTIN, S. KRUMSCHEID, AND F. NOBILE, *Complexity analysis of stochastic gradient methods for PDE-constrained optimal control problems with uncertain parameters*, ESAIM Math. Model. Numer. Anal., 55 (2021), pp. 1599–1633.

[27] L. NG AND K. WILLCOX, *Multifidelity approaches for optimization under uncertainty*, Internat. J. Numer. Methods Engrg., 100 (2014), pp. 746–772.

[28] B. ØKSENDAL, *Stochastic Differential Equations*, Springer, New York, 2003.

[29] C. PHELPS, J. O. ROYSET, AND Q. GONG, *Optimal control of uncertain systems using sample average approximations*, SIAM J. Control Optim., 54 (2016), pp. 1–29, https://doi.org/10.1137/140983161.

[30] A. QUARTERONI AND S. QUARTERONI, *Numerical Models for Differential Problems*, 3rd ed., Mathematics and Statistics 16, Springer, Cham, 2017.

[31] R. T. ROCKAFELLAR AND J. O. ROYSET, *Measures of residual risk with connections to regression, risk tracking, surrogate models, and ambiguity*, SIAM J. Optim., 25 (2015), pp. 1179–1208, https://doi.org/10.1137/151003271.

[32] R. T. ROCKAFELLAR AND S. URYASEV, *Optimization of conditional value-at-risk*, J. Risk, 2 (2000), pp. 493–517.

[33] R. T. ROCKAFELLAR AND S. URYASEV, *The fundamental risk quadrangle in risk management, optimization and statistical estimation*, Surv. Oper. Res. Manag. Sci., 18 (2013), pp. 33–53.

[34] R. T. ROCKAFELLAR AND J. O. ROYSET, *On buffered failure probability in design and optimization of structures*, Reliab. Eng. Syst. Saf., 95 (2010), pp. 499–510.

[35] W. RÖMISCH AND T. M. SUROWIEC, *Asymptotic Properties of Monte Carlo Methods in Elliptic PDE-Constrained Optimization under Uncertainty*, arXiv:2106.06347, 2021.

[36] R. T. ROCKAFELLAR AND R. J.-B. WETS, *Variational Analysis*, Springer, New York, 1998.

[37] J. O. ROYSET AND R. J.-B. WETS, *Variational analysis of constrained M-estimators*, Ann. Statist., 48 (2020), pp. 2759–2790.

[38] J. O. ROYSET AND R. J.-B. WETS, *An Optimization Primer*, Springer, New York, 2021.

[39] C. SCHWAB AND R. A. TODOR, *Karhunen-Loève approximation of random fields by generalized fast multipole methods*, J. Comput. Phys., 217 (2006), pp. 100–122.

[40] A. SCHWARTZ AND E. POLAK, *Consistent approximations for optimal control problems based on Runge–Kutta integration*, SIAM J. Control Optim., 34 (1996), pp. 1235–1269, https://doi.org/10.1137/S0363012994267352.

[41] F. TRÖLTZSCH, *Optimal Control of Partial Differential Equations: Theory, Methods, and Applications*, Grad. Stud. Math. 112, AMS, Providence, RI, 2010.

[42] F. WECHSUNG, A. GIULIANI, M. LANDREMAN, A. CERFON, AND G. STADLER, *Single-stage gradient-based stellarator coil design: Stochastic optimization*, Nucl. Fusion, 62 (2022), 076034.

[43] H. YANG AND M. GUNZBURGER, *Algorithms and analyses for stochastic optimization for turbofan noise reduction using parallel reduced-order modeling*, Comput. Methods Appl. Mech. Engrg., 319 (2017), pp. 217–239.