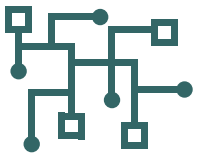

ISE

Industrial and
Systems Engineering

State-Constrained Optimization Problems Under Uncertainty: A Tensor Train Approach

HARBIR ANTIL[1], SERGEY DOLGOV[2], AND AKWUM ONWUNTA[3]

[1]The Center for Mathematics and Artificial Intelligence (CMAI) and Department of Mathematical Sciences, George Mason University, Fairfax, VA, USA

[2]Department of Mathematical Sciences, University of Bath, Bath, UK

[3]Department of Industrial and Systems Engineering, Lehigh University, Bethlehem, PA, USA

ISE Technical Report 23T-011

LEHIGH
UNIVERSITY

STATE-CONSTRAINED OPTIMIZATION PROBLEMS UNDER UNCERTAINTY: A TENSOR TRAIN APPROACH

HARBIR ANTIL, SERGEY DOLGOV, AND AKWUM ONWUNTA

Abstract. We propose an algorithm to solve optimization problems constrained by partial (ordinary) differential equations under uncertainty, with almost sure constraints on the state variable. To alleviate the computational burden of high-dimensional random variables, we approximate all random fields by the tensor-train decomposition. To enable efficient tensor-train approximation of the state constraints, the latter are handled using the Moreau-Yosida penalty, with an additional smoothing of the positive part (plus/ReLU) function by a softplus function. We derive theoretical bounds on the constraint violation in terms of the Moreau-Yosida regularization parameter and smoothing width of the softplus function. This result also proposes a practical recipe for selecting these two parameters. When the optimization problem is strongly convex, we establish strong convergence of the regularized solution to the optimal control. We develop a second order Newton type method with a fast matrix-free action of the approximate Hessian to solve the smoothed Moreau-Yosida problem. This algorithm is tested on benchmark elliptic problems with random coefficients, optimization problems constrained by random elliptic variational inequalities, and a real-world epidemiological model with 20 random variables. These examples demonstrate mild (at most polynomial) scaling with respect to the dimension and regularization parameters.

1. Introduction

Over last two decades optimization problems constrained by physical laws, such as partial (ordinary) differential equations (PDEs/ODEs), have emerged as a prominent research area. This is fueled by many applications in science and engineering, such as controlling pathogen propagation in built environment [26, 25], shape and topology optimization [36, 28], optimal strategies to predict shutdowns due to pandemics [11]. The optimization variables consist of state (y) and control/design (u). However, often due to noisy measurements and ambiguous models due to incomplete physics, the underlying physical laws contain uncertainty. This has led to significant theoretical and algorithmic developments in the area of optimization problems constrained by physical laws under uncertainty. See for instance [23, 4, 14, 3] and the references therein. These papers focus on problems with control constraints.

The literature on state-constrained optimization problems under uncertainty is scarce. For instance, [12, 17] use probability constraints, and [15, 13, 16] consider almost surely type constraints. It is well-known that even in the deterministic setting, the state constrained problems are highly challenging. One of the fundamental difficulties is that the

2020 *Mathematics Subject Classification.* 49J55, 93E20, 49K20, 49K45, 90C15, 65D15, 15A69, 15A23 .

Key words and phrases. almost surely constraints, state constraints, risk neutral, tensor train, reduced space, preconditioner, variational inequality.

HA is partially supported by NSF grant DMS-2110263 and the AirForce Office of Scientific Research under Award NO: FA9550-22-1-0248.

state constraints are imposed in the sense of continuous functions. As a result, the Lagrange multipliers corresponding to those constraints are Radon measures that exhibit low regularity [6]. The situation is much more delicate in the stochastic setting. We refer to the aforementioned references for a detailed discussion on this topic. Motivated by the deterministic setting, [13] introduces a Moreau-Yosida based approximation scheme to solve the state-constrained optimization problems when the PDE constraints are given by an elliptic equation with random coefficients. Further extensions of this work are considered in [1, 16]. However, all of these papers approximate expectations of random fields by Monte-Carlo-type methods, which may converge slowly.

In [3], we introduced an algorithm (TTRISK) based on the tensor train (TT) decomposition [30] to solve risk-averse optimization problems with control constraints, and the conditional value-at-risk (CVaR) [32] risk measure. We demonstrated that the extra computational cost due to the uncertainty can scale proportionally to $\text{error}^{-0.5}$ when the TT approximation is used, in contrast to a error^{-2} scaling of Monte Carlo quadratures. In the current paper, we continue this program and develop a TT based algorithm for state-constrained optimization problems. For simplicity of presentation, we only consider the risk-neutral setting, i.e., the objective function is given by the expected value of a quantity of interest. Similarly to [13, 16], we tackle the state constrains using Moreau-Yosida based relaxation with a softplus smoothing. The main contributions of this paper are listed next:

(i) We consider an ε-softplus regularization of the positive part function $(\cdot)_+ = \max\{\cdot, 0\}$ and derive a probabilistic estimate of state constraint violation in terms of Moreau-Yosida regularization parameter γ and ε. In particular, we show that selecting $\varepsilon \propto \gamma^{-1/2}$ ensures the convergence of the constraint violation with a rate $\gamma^{-1/2}$. This result is motivated by [13, Prop. 2]. Notice that the ε-smoothing is carried out because the irregular function $(\cdot)_+$ may lack an efficient TT decomposition.
(ii) When the optimization problem is strongly convex, we establish strong convergence of the regularized solution to the optimal control. Our final results can be seen as generalizations of the results in deterministic setting.
(iii) We derive a second order Newton type method to solve the regularized problem with a fast matrix-free action of the approximate Hessian.
(iv) We test the proposed method on elliptic equations in one and two physical dimensions and random coefficients, as well as an ODE example (motivated by a realistic application) with 20 random variables, and show that the algorithm is free from the curse of dimensionality.
(v) The proposed approach has been also successfully applied to an example where the PDE constraint is given by an elliptic variational inequality.

Outline: The remainder of the paper is organized as follows. In Section 2, we provide a rigorous mathematical formulation of the problem under consideration. Section 3 is devoted to the Moreau-Yosida approximation, derivation of the second order Newton method and approximation error estimates due to the Moreau-Yosida approximation. In Section 4, we provide a brief description of the TT format. This is followed by practical aspects of Moreau-Yosida approximations in Section 5. Finally, in Section 6, we provide a series of numerical experiments. At first, we consider an optimization problem with an elliptic PDE in one spatial dimension as constraints. This is followed by a two-dimensional case. After these

benchmarks, an optimization problem with an elliptic variational inequality as constraint is considered in Section 6.3. The numerical experiments conclude with a realistic ODE example for designing optimal lockdown strategies in Section 6.4.

2. Problem Formulation

Let $(\Omega, \mathcal{F}, \mathbb{P})$ denote a complete probability space, where Ω represents the sample space, $\mathcal{F}$ is the Borel σ-algebra of events on the power set of Ω, and $\mathbb{P} : \Omega \to [0, 1]$ is an appropriate probability measure. We denote by $\mathbb{E}[\cdot]$ the expectation with respect to $\mathbb{P}$. Let $\mathcal{U}$ be a real deterministic reflexive Banach space of optimization variables (control or design) defined on an open, bounded and connected set $D \subset \mathbb{R}^n$ with Lipschitz boundary. We denote by $\|\cdot\|_{\mathcal{U}}$ the norm on $\mathcal{U}$, and the duality pairing between $\mathcal{U}$ and $\mathcal{U}^*$ as $\langle\cdot,\cdot\rangle_{\mathcal{U}^*,\mathcal{U}}$. Let $\mathcal{Y} = L^2(\Omega, \mathcal{F}, \mathbb{P}; \hat{\mathcal{Y}})$ and $\mathcal{Z} = L^2(\Omega, \mathcal{F}, \mathbb{P}; \hat{\mathcal{Z}})$ be Bochner spaces of random fields, based on deterministic Banach spaces $\hat{\mathcal{Y}} \hookrightarrow L^2(D) \hookrightarrow \hat{\mathcal{Y}}^*$ and $\hat{\mathcal{Z}}$, with corresponding norms and duality pairings

$$\|y\|_{\mathcal{Y}}^2 = \mathbb{E}[\|y(\omega)\|_{\hat{\mathcal{Y}}}^2], \qquad \langle y, v\rangle_{\mathcal{Y}^*,\mathcal{Y}} = \mathbb{E}\left[\langle y(\omega), v(\omega)\rangle_{\hat{\mathcal{Y}}^*,\hat{\mathcal{Y}}}\right],$$

and similarly for $\mathcal{Z}$. Let $\mathcal{U}_{ad} \subseteq \mathcal{U}$ be a closed convex nonempty subset and let $c : \mathcal{Y}\times\mathcal{U}_{ad}\times\Omega \to \mathcal{Z}$ denote, e.g., a partial differential operator, then consider the equality constraint

$$c(y, u; \omega) = 0, \quad \text{in } \mathcal{Z}, \quad \text{a.s. } \omega \in \Omega,$$

where a.s. indicates "almost surely" with respect to the probability measure $\mathbb{P}$.

In this paper, we consider the optimization problems of the form

$$\min_{y,u} \mathcal{R}[J(y, u; \omega)] \tag{2.1}$$

$$\text{s.t} \quad c(y, u; \omega) = 0, \quad \text{in } \mathcal{Z}, \quad \text{a.s. } \omega \in \Omega, \tag{2.2}$$

where $\mathcal{R}$ represents the risk measure and $\mathcal{R}[J(y, u; \omega)]$ is a deterministic cost function. More precisely, we will focus on the so-called risk-neutral formulation; that is, $\mathcal{R}$ is simply the expectation, denoted by $\mathbb{E}$. We are particularly interested in the case in which the state variable y is constrained by a random variable:

$$y \le y_{\max}(\omega) \quad \text{a.s.}, \tag{2.3}$$

where we assume that $y_{\max} \in \mathcal{Y}$.

In what follows, we disuss the Moreau-Yosida approximation for (2.1)-(2.3) and derive a Newton type method. Throughout the paper, without explicitly stating, we will make use of the following assumption.

Assumption 2.1 (unique forward solution)**.** *There exists an injective operator $S(\omega) : \mathcal{U}_{ad} \to \mathcal{Y}$ (maybe nonlinear) such that $c(S(\omega)u, u; \omega) = 0 \ \forall u \in \mathcal{U}_{ad}$ a.s.*

This allows us to define the reduced-space cost function

$$j(u) := \mathcal{R}[J(S(\omega)u, u; \omega)]. \tag{2.4}$$

The resulting reduced optimization problem is given by

$$\begin{aligned} &\min_{u\in\mathcal{U}_{ad}} j(u) \\ &\text{s.t} \quad y \le y_{\max}(\omega) \quad \text{a.s.} \end{aligned} \tag{2.5}$$

3. Smoothed Moreau-Yosida approximation

Solving (2.5) with state constraints involve computation of the indicator function of an active set and/or Lagrange multiplier as a random field that is nonnegative on a complicated high-dimensional domain. This may be difficult for many function approximation methods, especially for tensor decompositions that are considered in this paper. We tackle this difficulty by first turning the constrained optimization problem (2.5) into an unconstrained optimization problem with the Moreau-Yosida penalty, and further by smoothing the indicator function in the penalty term.

The classical Moreau-Yosida problem reads, with $\gamma \geq 0$ denoting the regularization parameter,

$$\min_{u \in \mathcal{U}_{ad}} j^{\gamma}(u), \quad \text{where} \quad j^{\gamma}(u) := j(u) + \frac{\gamma}{2}\mathbb{E}\left[\left\|(Su - y_{\max}(\xi))_+\right\|^2_{L^2(D)}\right], \tag{3.1}$$

where the so-called *positive part* or *ReLU* function $(\cdot)_+$ reads $(s)_+ = s$ if $s \geq 0$ and 0, otherwise. Here, we have removed the need to optimize the Lagrange multiplier (corresponding to the inequality constraints) over the nonnegative cone, but the function approximation of a nonsmooth high-dimensional random field $(Su - y_{\max}(\xi))_+$ (and derivatives thereof) may be still inefficient.

For this reason, we replace the *ReLU* function in the penalty term by a smoothed version. In this paper, we use the *softplus* function

$$g_\varepsilon(s) = \varepsilon \cdot \log(1 + \exp(s/\varepsilon)) \in C^\infty(\mathbb{R}), \qquad g_0(s) = \lim_{\varepsilon \to 0} g_\varepsilon(x) = (s)_+, \tag{3.2}$$

although other (e.g. piecewise polynomial) functions are also possible [24, 1]. Now, the cost function becomes

$$j^{\gamma,\varepsilon}(u) := j(u) + \frac{\gamma}{2}\mathbb{E}\left[\left\|g_\varepsilon(Su - y_{\max})\right\|^2_{L^2(D)}\right]. \tag{3.3}$$

3.1. Discretization and Derivatives of the Cost.

In practice, the operator S involves the solution of a differential equation, which needs to be discretized (using e.g. Finite Element methods and/or time integration schemes). For a given mesh parameter $h > 0$, we introduce the discretized (maybe nonlinear) operator $\mathbf{S}_h(\omega) : \mathcal{U}_{ad} \to \mathbb{R}^{n_y}$, where n_y is the total number of degrees of freedom in the discrete solution. We denote the induced Bochner space $\mathcal{Y}_h \cong L^2_h(\Omega, D) := L^2(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^{n_y})$. The L^2-norm can be written as an expectation of a vector quadratic form,

$$\|\mathbf{y}\|^2_{L^2_h(\Omega,D)} = \mathbb{E}\left[\mathbf{y}(\omega)^\top \mathbf{M}\mathbf{y}(\omega)\right], \qquad \forall \mathbf{y} \in L^2_h(\Omega, D),$$

where $\mathbf{M} = \mathbf{M}^\top > 0 \in \mathbb{R}^{n_y \times n_y}$ is a mass matrix. The discretized problem cost is denoted by $j^h(u) \approx j(u)$, and the discretized constraint is $\mathbf{y}^h_{\max} \in \mathcal{Y}_h$. Now, the semi-discretized Moreau-Yosida cost function (3.3) becomes

$$j^{\gamma,\varepsilon,h}(u) := j^h(u) + \frac{\gamma}{2}\mathbb{E}\left[\|g_\varepsilon(\mathbf{S}_h u - \mathbf{y}^h_{\max})\|^2_{\mathbf{M}}\right]. \tag{3.4}$$

To derive a Newton type method, we compute the expressions of gradient and Hessian:

$$\nabla_u j^{\gamma,\varepsilon,h} = \nabla_u j^h + \gamma \mathbb{E}\left[\mathbf{S}_h^* \cdot \operatorname{diag}(g_\varepsilon'(\mathbf{S}_h u - \mathbf{y}_{\max}^h)) \cdot \mathbf{M} g_\varepsilon(\mathbf{S}_h u - \mathbf{y}_{\max}^h)\right], \tag{3.5}$$

$$\nabla_{uu} j^{\gamma,\varepsilon,h} = \nabla_{uu}^2 j^h + \gamma \mathbb{E}\left[\mathbf{S}_h^* \cdot \operatorname{diag}(g_\varepsilon') \mathbf{M} \operatorname{diag}(g_\varepsilon') \cdot \mathbf{S}_h'\right] \tag{3.6}$$

$$+ \gamma \mathbb{E}\left[\mathbf{S}_h^* \cdot (\operatorname{tendiag}(g_\varepsilon'') \times_3 (\mathbf{M} g_\varepsilon)) \cdot \mathbf{S}_h'\right] \tag{3.7}$$

$$+ \gamma \mathbb{E}\left[\nabla_u \mathbf{S}_h^* \times_3 (\operatorname{diag}(g_\varepsilon'(\mathbf{S}_h u - \mathbf{y}_{\max}^h)) \cdot \mathbf{M} g_\varepsilon(\mathbf{S}_h u - \mathbf{y}_{\max}^h))\right], \tag{3.8}$$

where tendiag$(\cdot)$ is producing a 3-dimensional tensor out of vector by putting the vector elements along the diagonal, and zero elements otherwise, and $\times_3$ is the tensor-vector contraction product over the 3d mode of the tensor. If $\mathbf{S}_h$ is a nonlinear operator, $\mathbf{S}_h' = \nabla_u \mathbf{S}_h(u)$ denotes the gradient of an image of u, and $\mathbf{S}_h^*$ is the adjoint of $\mathbf{S}_h'$.

3.2. **Matrix-free Fixed Point Gauss-Newton Hessian.** The exact assembly of all terms of the Hessian (3.6)–(3.8) can be too computationally expensive, since this involves dense tensor-valued random fields (such as $\nabla_u \mathbf{S}_h^*$). To simplify the computations, we can firstly omit the terms (3.7) and (3.8) which contain order-3 tensors. Secondly, we can replace the exact expectation by a fixed-point evaluation. Rewriting (2.1) using Assumption 2.1 we can define $J(u;\omega) = J(S(\omega)u, u;\omega)$ and its discretized version $J^h(u;\omega) = J(\mathbf{S}_h(\omega)u, u;\omega)$. The Hessian of j^h can then be written as

$$\nabla_{uu}^2 j^h = \mathbb{E}\left[\nabla_{uu}^2 J^h(u;\omega)\right].$$

For practical computations, it is convenient to parametrize all random fields with independent identically distributed (i.i.d.) random variables with a known probability density function. Those variables can then be sampled independently, and an expectation can be computed simply by quadrature. Therefore, we will use the following assumption.

Assumption 3.1 (finite noise)**.** *There exists a d-dimensional random vector $\xi(\omega) \in \mathbb{R}^d$ with a product probability density function $\pi(\xi) = \pi(\xi_1) \cdots \pi(\xi_d)$, such that any random field $y \in \mathcal{Y}$ can be expressed as a function of ξ, $y(\omega) = y(\xi(\omega))$ a.s., and*

$$\mathbb{E}[y] = \int_{\mathbb{R}^d} y(\xi)\pi(\xi)d\xi.$$

In particular, the vector ξ can often be derived from a parametrization of the forward solution operator $\mathbf{S}_h(\omega) = \mathbf{S}_h(\xi(\omega))$, and/or the constraint $\mathbf{y}_{\max}^h(\omega) = \mathbf{y}_{\max}^h(\xi(\omega))$.

Example 3.2. *Let $y = S(\nu(\omega))u$ be the resolution of an elliptic PDE*

$$-\nabla(\kappa(x;\nu(\omega))\nabla y) = u,$$

where the diffusivity

$$\kappa(x;\nu(\omega)) = \kappa_0(x) + \sum_{k=1}^{p} \psi_k(x)\nu_k(\omega)$$

and the constraint

$$y_{\max}(x;\eta(\omega)) = y_0(x) + \sum_{k=1}^{q} \phi_k(x)\eta_k(\omega)$$

are given by Karhunen-Loeve expansions (see e.g., [27]), where ν and η are independent random variables. Then, we can define $\xi = (\nu_1, \ldots, \nu_p, \eta_1, \ldots, \eta_q)$.

Now we can replace $\nabla^2_{uu} j^h = \mathbb{E}[\nabla^2_{uu} J^h(u;\xi)]$ by

$$\tilde{\nabla}^2_{uu} j^h = \nabla^2_{uu} J^h(u; \mathbb{E}[\xi]).$$

This is exact if $\nabla^2_{uu} J^h$ is linear in ξ, but we can take it as an approximation in the general case too. Now to apply $\tilde{\nabla}^2_{uu} j^h$ to a vector we just need to apply one deterministic $\nabla^2_{uu} J^h(u; \mathbb{E}[\xi])$, which involves solving one forward, one adjoint, and two linear sensitivity (of state and adjoint) deterministic problems in the most general setting [4, Ch. 1, Algo. 2].

Similarly we approximate the second term in (3.6) by

$$\gamma \mathbf{S}_h^*(\xi_*)\mathbf{M}\mathbf{S}_h'(\xi_*),$$

where

$$\xi_* = \frac{\mathbb{E}\left[\xi \cdot \mathbf{1}^\top g_\varepsilon'(\mathbf{S}_h u - \mathbf{y}^h_{\max}(\xi))\right]}{\mathbb{E}\left[\mathbf{1}^\top g_\varepsilon'(\mathbf{S}_h u - \mathbf{y}^h_{\max}(\xi))\right]}$$

is the mean of the random variable with respect to the probability density $\pi_{g'} \propto \pi \cdot (\mathbf{1}^\top g_\varepsilon'(\mathbf{S}_h u - \mathbf{y}^h_{\max}))$, and $\mathbf{1} \in \mathbb{R}^{n_y}$ is the constant vector, averaging the spatial components. Note that $\mathbf{1}^\top g_\varepsilon'(\mathbf{S}_h u - \mathbf{y}^h_{\max})$ is a nonnegative function bounded by n_y, so $\pi \mathbf{1}^\top g_\varepsilon'(\mathbf{S}_h u - \mathbf{y}^h_{\max})$ is nonnegative and normalizable, and $\pi_{g'}$ is indeed a probability density.

Finally, we obtain a deterministic approximate Hessian

$$\tilde{\mathbf{H}} = \nabla^2_{uu} J^h(u; \mathbb{E}[\xi]) + \gamma \mathbf{S}_h^*(\xi_*)\mathbf{M}\mathbf{S}_h'(\xi_*), \tag{3.9}$$

which can be applied to a vector by solving 2 forward, 2 adjoint, and 2 sensitivity problems.

3.3. **Probability of the Constraint Violation.** In the rest of this section, we prove certain properties about the quality of the solution of the smoothed problem (3.3) with respect to the constraint, and the exact solution of (2.1)–(2.3). This needs a few properties of the softplus smoothing function.

Lemma 3.3. *For any $\varepsilon \geq 0$, the softplus function (3.2) satifies: $g_\varepsilon(s) \geq (s)_+$ for any $s \in \mathbb{R}$, $g_\varepsilon'(s) \geq 0.5$ for $s \geq 0$, and $g_\varepsilon'(s) \leq 0.5$ for $s \leq 0$.*

Proof. Using the monotonicity of the logarithm,

$$g_\varepsilon(s) = \varepsilon \log\left(1 + \exp(s/\varepsilon)\right) \geq \begin{cases} \varepsilon \log\left(\exp(s/\varepsilon)\right) = s = (s)_+, & s \geq 0, \\ 0 = (s)_+, & s < 0. \end{cases}$$

The remaining inequalities follow simply from the monotonicity of the sigmoid function $g_\varepsilon'(s) = 1/(1 + \exp(-s/\varepsilon))$ and that $g_\varepsilon'(0) = 0.5$. □

Theorem 3.4. *Let $u^{\gamma,\varepsilon}$ be a minimizer of (3.3), and assume that $j(u) \geq 0$ for any $u \in \mathcal{U}_{ad}$. Then for any $\delta > 0$, we have*

$$\mathbb{P}\left[\|(S(\omega)u^{\gamma,\varepsilon} - y_{\max}(\omega))_+\|^2_{L^2(D)} > \delta\right] \leq \frac{C_1 + C_2\gamma\varepsilon^2}{\gamma\delta},$$

where $C_1 = 2j(u_*)$, $C_2 = \log^2 2 \cdot \|1\|^2_{L^2(D)}$, and u_* is a minimizer of (2.1)–(2.3).

Remark 3.5. *This motivates the condition $\varepsilon \lesssim 1/\sqrt{\gamma}$ to overcome the effect of smoothing.*

Proof. Using Markov's inequality, we obtain

$$\mathbb{P}\left[\|(Su^{\gamma,\varepsilon}-y_{\max}(\omega))_+\|^2_{L^2(D)}>\delta\right]\le\frac{\mathbb{E}\left[\left\|(Su^{\gamma,\varepsilon}-y_{\max}(\omega))_+\right\|^2_{L^2(D)}\right]}{\delta}\le\frac{\mathbb{E}\left[\left\|g_\varepsilon(Su^{\gamma,\varepsilon}-y_{\max}(\omega))\right\|^2_{L^2(D)}\right]}{\delta},$$

where in the second inequality we used Lemma 3.3. Since $u^{\gamma,\varepsilon}$ minimizes (3.3), it holds

$$j(u^{\gamma,\varepsilon})+\frac{\gamma}{2}\mathbb{E}[\|g_\varepsilon(Su^{\gamma,\varepsilon}-y_{\max}(\omega))\|^2_{L^2(D)}]\le j(u_*)+\frac{\gamma}{2}\mathbb{E}[\|g_\varepsilon(Su_*-y_{\max}(\omega))\|^2_{L^2(D)}]$$

for any $u_*\in\mathcal{U}_{ad}$ such as the minimizer of (2.1) constrained to (2.3). Dividing by $\gamma/2$ and neglecting $j(u^{\gamma,\varepsilon})\ge 0$, we get

$$\mathbb{E}[\|g_\varepsilon(Su^{\gamma,\varepsilon}-y_{\max}(\omega))\|^2_{L^2(D)}]\le\frac{C_1}{\gamma}+\mathbb{E}[\|g_\varepsilon(Su_*-y_{\max}(\omega))\|^2_{L^2(D)}].$$

For the latter term, (2.3) implies $Su_*-y_{\max}(\omega)\le 0$ a.s., and due to monotonicity of g_ε,

$$g_\varepsilon(Su_*-y_{\max}(\omega))\le g_\varepsilon(0)=\varepsilon\cdot\log 2\qquad\text{a.s.}$$

Taking this upper bound out of the expectation and norm, we obtain

$$\mathbb{E}[\|g_\varepsilon(Su^{\gamma,\varepsilon}-y_{\max}(\omega))\|^2_{L^2(D)}]\le\frac{C_1}{\gamma}+\varepsilon^2\cdot\log^2 2\cdot\mathbb{E}[\|1\|^2_{L^2(D)}]=\frac{C_1}{\gamma}+\varepsilon^2\cdot\log^2 2\cdot\|1\|^2_{L^2(D)},\quad(3.10)$$

and the estimate on probability follows by the Markov's inequality. □

3.4. **Strong Convergence with Strongly Convex Cost.** To prove the strong convergence of the minimizer of (3.3) to the minimizer of (2.1)–(2.3) we need further assumptions on the cost and smoothing functions.

Assumption 3.6 (Bounded derivative of the cost)**.** *There exists $L<\infty$ such that*

$$\|j'(u)\|_{\mathcal{U}^*}\le L\qquad\forall u\in\mathcal{U}_{ad}.$$

Assumption 3.7 (α-strong convexity of the cost)**.** *There exists $\alpha>0$ such that*

$$\langle j'(u)-j'(v),u-v\rangle_{\mathcal{U}^*,\mathcal{U}}\ge\alpha\|u-v\|^2_{\mathcal{U}},\qquad\forall u,v\in\mathcal{U}_{ad}.$$

Assumption 3.8 (Smoothing function)**.** *The smoothing function g_ε possesses the following properties*

$$\begin{aligned}&g'_\varepsilon(s)\ge 0.5,\quad g_\varepsilon(s)\ge s,\quad \textit{for}\quad s\ge 0,\\&g'_\varepsilon(s)\le 0.5,\qquad\qquad\qquad\ \ \textit{for}\quad s\le 0,\end{aligned}\quad(3.11)$$

and either:

$$g_\varepsilon(s)s\ge-\eta_{max}(\varepsilon),\quad\textit{for}\quad s\le 0,\quad(3.12)$$

or, for any random field $y(\omega)\in\mathcal{Y}$ such that $y(\omega)\le 0$ a.s.,

$$\langle y,g_\varepsilon(y)\rangle_{\mathcal{Y}^*,\mathcal{Y}}\ge-\eta_{int}(\varepsilon),\quad(3.13)$$

where $\eta_{max}(\varepsilon),\eta_{int}(\varepsilon)\ge 0$, $\forall\varepsilon>0$, $\eta_{max}(\varepsilon),\eta_{int}(\varepsilon)\to 0$ as $\varepsilon\to 0$.

Notice that all the conditions in (3.11) are satisfied by the *softplus* function (3.2) (see Lemma 3.3). We only need to check (3.12) or alternatively (3.13).

Conjecture 3.9. *Our numerical experiments demonstrate that for the* softplus *function* (3.2) *it holds* $\eta_{max}(\varepsilon) = \mathcal{O}(\varepsilon^2)$ *and* $\eta_{int}(\varepsilon) = \mathcal{O}(\varepsilon^3)$*, although we are only able to prove the latter estimate under specific conditions (Lemma 3.11 and Theorem 3.12).*

Now we are able to prove the strong convergence of the smoothed optimal control.

Theorem 3.10. *Under Assumptions 2.1 and 3.6–3.8, linear operator* S*, and* $\varepsilon = \varepsilon_\gamma$ *dependent on* γ *in such a way that*

$$\gamma \min\{\eta_{max}(\varepsilon_\gamma), \eta_{int}(\varepsilon_\gamma)\} \to 0, \quad \text{as} \quad \gamma \to \infty,$$

and $\langle f, f\rangle_{\mathcal{Y}^*,\mathcal{Y}} = \|f\|^2_{L^2(\Omega,D)}$ *for any* $f \in \mathcal{Y}$*, the minimizer* u_γ *of* (3.3) *converges to the solution* u_* *of the exact problem* (2.1)–(2.3)*,*

$$\alpha\|u_\gamma - u_*\|^2_{\mathcal{U}} + \frac{\gamma}{2}\|(Su_\gamma - y_{\max})_+\|^2_{L^2(\Omega,D)} \to 0, \qquad \gamma \to \infty.$$

Proof. The optimality condition for the smoothed problem, $\langle \nabla_u j^{\gamma,\varepsilon}(u_\gamma), v - u_\gamma\rangle_{\mathcal{U}^*,\mathcal{U}} \geq 0$, $\forall v \in \mathcal{U}_{ad}$, can be expanded by introducing an auxiliary variable λ_γ to match the gradient of the Moreau-Yosida term:

$$\langle j'(u_\gamma) + S^*\lambda_\gamma, v - u_\gamma\rangle_{\mathcal{U}^*,\mathcal{U}} \geq 0, \tag{3.14}$$

$$\gamma g'_\varepsilon(Su_\gamma - y_{\max}) g_\varepsilon(Su_\gamma - y_{\max}) = \lambda_\gamma. \tag{3.15}$$

In turn, the KKT conditions for the original problem read

$$\langle j'(u_*) + S^*\lambda_*, v - u_*\rangle_{\mathcal{U}^*,\mathcal{U}} \geq 0 \qquad \forall v \in \mathcal{U}_{ad} \tag{3.16}$$

$$\lambda_* \geq 0$$

$$Su_* - y_{\max} \leq 0$$

$$\langle \lambda_*, Su_* - y_{\max}\rangle_{\mathcal{Y}^*,\mathcal{Y}} = 0. \tag{3.17}$$

Adding (3.16) with $v = u_\gamma$ to (3.14) with $v = u_*$, and casting S^* onto another side of the duality pairing, we get

$$\begin{aligned} 0 &\geq \langle j'(u_\gamma) + S^*\lambda_\gamma - j'(u_*) - S^*\lambda_*, u_\gamma - u_*\rangle_{\mathcal{U}^*,\mathcal{U}} \\ &= \langle j'(u_\gamma) - j'(u_*), u_\gamma - u_*\rangle_{\mathcal{U}^*,\mathcal{U}} + \langle \lambda_\gamma, Su_\gamma - Su_*\rangle_{\mathcal{Y}^*,\mathcal{Y}} + \langle j'(u_*), u_\gamma - u_*\rangle_{\mathcal{U}^*,\mathcal{U}}. \end{aligned} \tag{3.18}$$

Due to the strong convexity, (3.18), and Assumption 3.6 we arrive at

$$\alpha\|u_\gamma - u_*\|^2_{\mathcal{U}} + \langle \lambda_\gamma, Su_\gamma - Su_*\rangle_{\mathcal{Y}^*,\mathcal{Y}} \leq \langle j'(u_*), u_* - u_\gamma\rangle_{\mathcal{U}^*,\mathcal{U}} \leq \|j'(u_*)\|_{\mathcal{U}^*}\|u_* - u_\gamma\|_{\mathcal{U}}. \tag{3.19}$$

The second term on the left hand side can be bounded as follows. Using the fact that $y_{\max} - Su_* \geq 0$ a.s. and the definition of λ_γ, we obtain that

$$\begin{aligned} \langle \lambda_\gamma, Su_\gamma - Su_*\rangle_{\mathcal{Y}^*,\mathcal{Y}} &= \langle \lambda_\gamma, (Su_\gamma - y_{\max}) + (y_{\max} - Su_*)\rangle_{\mathcal{Y}^*,\mathcal{Y}} \\ &\geq \langle \lambda_\gamma, Su_\gamma - y_{\max}\rangle_{\mathcal{Y}^*,\mathcal{Y}} \\ &= \gamma\langle g'_\varepsilon(Su_\gamma - y_{\max}) g_\varepsilon(Su_\gamma - y_{\max}), Su_\gamma - y_{\max}\rangle_{\mathcal{Y}^*,\mathcal{Y}} \\ &= \gamma\langle g'_\varepsilon(Su_\gamma - y_{\max})(Su_\gamma - y_{\max}), g_\varepsilon(Su_\gamma - y_{\max})\rangle_{\mathcal{Y}^*,\mathcal{Y}} \\ &= \gamma\langle g'_\varepsilon(Su_\gamma - y_{\max})(Su_\gamma - y_{\max})_+, g_\varepsilon(Su_\gamma - y_{\max})\rangle_{\mathcal{Y}^*,\mathcal{Y}} \\ &\quad + \gamma\langle g'_\varepsilon(Su_\gamma - y_{\max})(Su_\gamma - y_{\max})_-, g_\varepsilon(Su_\gamma - y_{\max})\rangle_{\mathcal{Y}^*,\mathcal{Y}}, \end{aligned} \tag{3.20}$$

where we have split $Su_\gamma - y_{\max}$ into positive and negative parts, with $(s)_- = \min(s, 0)$ denoting the negative part. Next using Assumption 3.8 in (3.20), we readily obtain that

$$\begin{aligned}\langle \lambda_\gamma, Su_\gamma - Su_*\rangle_{\mathcal{Y}^*,\mathcal{Y}} &\geq \gamma\langle 0.5(Su_\gamma - y_{\max})_+, (Su_\gamma - y_{\max})_+\rangle_{\mathcal{Y}^*,\mathcal{Y}} \\ &\quad + \gamma\langle 0.5(Su_\gamma - y_{\max})_-, g_\varepsilon(Su_\gamma - y_{\max})\rangle_{\mathcal{Y}^*,\mathcal{Y}} \qquad (3.21)\\ &\geq \gamma\left[0.5\|(Su_\gamma - y_{\max})_+\|^2_{L^2(\Omega,D)} - 0.5\eta_{int}(\varepsilon)\right]. \qquad (3.22)\end{aligned}$$

Alternatively, we can bound (3.21) using (3.12) to arrive at

$$\langle \lambda_\gamma, Su_\gamma - Su_*\rangle_{\mathcal{Y}^*,\mathcal{Y}} \geq \gamma\left[0.5\|(Su_\gamma - y_{\max})_+\|^2_{L^2(\Omega,D)} - 0.5\eta_{max}(\varepsilon)\|1\|^2_{L^2(\Omega,D)}\right].$$

In either case, (3.19) implies that u_γ is bounded in $\mathcal{U}_{ad}$. Therefore, there exists a weakly converging subsequence $u_\gamma \rightharpoonup \hat{u}$ in $\mathcal{U}$ as $\gamma \to \infty$. Since, $\mathcal{U}_{ad}$ is closed convex, therefore $\hat{u} \in \mathcal{U}_{ad}$. If $\varepsilon = \varepsilon_\gamma \to 0$ as $\gamma \to \infty$, Assumption 3.8 (for both η_{max} and η_{int}) implies that $0.5\gamma\|(Su_\gamma - y_{\max})_+\|^2_{L^2(\Omega,D)}$ is bounded, which means $\|(Su_\gamma - y_{\max})_+\|^2_{L^2(\Omega,D)} \to 0$ as $\gamma \to \infty$. Since S is injective and linear, $\|(Su_\gamma - y_{\max})_+\|^2_{L^2(\Omega,D)}$ is continuous and convex, hence [38, Theorem 2.12]:

$$0 = \liminf_{\gamma\to\infty} \|(Su_\gamma - y_{\max})_+\|^2_{L^2(\Omega,D)} \geq \|(S\hat{u} - y_{\max})_+\|^2_{L^2(\Omega,D)}.$$

Since D is a connected domain of positive measure, this yields $|(S\hat{u} - y_{\max})_+| = 0$, that is, $S\hat{u} \leq y_{\max}$ a.s. Adding again (3.16) and (3.14) and using strong convexity of j, but keeping both λ_γ and λ_*, we get

$$\begin{aligned}\alpha\|u_\gamma - u_*\|^2_{\mathcal{U}} &\leq \langle \lambda_* - \lambda_\gamma, Su_\gamma - Su_*\rangle_{\mathcal{Y}^*,\mathcal{Y}} \qquad (3.23)\\ &\leq \langle \lambda_*, (Su_\gamma - y_{\max}) + (y_{\max} - Su_*)\rangle_{\mathcal{Y}^*,\mathcal{Y}} \qquad (3.24)\\ &\quad - \frac{\gamma}{2}\|(Su_\gamma - y_{\max})_+\|^2_{L^2(\Omega,D)} + \frac{\gamma}{2}\min\{\|1\|^2_{L^2(\Omega,D)}\eta_{\max}(\varepsilon_\gamma), \eta_{int}(\varepsilon_\gamma)\}, \qquad (3.25)\end{aligned}$$

where we used (3.22) with the negative sign. If $\gamma\eta_{max}(\varepsilon_\gamma) \to 0$ or $\gamma\eta_{int}(\varepsilon_\gamma) \to 0$, then

$$0 \leq \lim_{\gamma\to\infty}[\alpha\|u_\gamma - u_*\|^2_{\mathcal{U}}] \leq \lim_{\gamma\to\infty}\langle \lambda_*, Su_\gamma - y_{\max}\rangle_{\mathcal{Y}^*,\mathcal{Y}} = \langle \underbrace{\lambda_*}_{\geq 0}, \underbrace{S\hat{u} - y_{\max}}_{\leq 0}\rangle_{\mathcal{Y}^*,\mathcal{Y}} \leq 0 \qquad (3.26)$$

due to (3.17), so $u_\gamma \to u_*$, thereby completing the proof of the theorem. □

Lemma 3.11. *For the softplus function* (3.2) *it holds for any* $\varepsilon \geq 0$*:*

$$\int_{-\infty}^{0} s g_\varepsilon(s) ds \geq -\varepsilon^3.$$

Proof. The proof uses elementary calculus and is given in Appendix A. □

In order to search for a rate of convergence, we establish the following result:

Theorem 3.12. *Suppose Assumptions 2.1, 3.1 and 3.6–3.8 hold,* $\hat{\mathcal{Y}}$ *is a space of scalar functions, the operator* S *is linear, and* $|\partial(Su - y_{\max})/\partial\xi_1| \geq c > 0$ *a.s.* $\forall u \in \mathcal{U}_{ad}$*. Suppose that* $\langle f, g\rangle_{\hat{\mathcal{Y}}^*,\hat{\mathcal{Y}}} = \int_D f(x)g(x)dx$ $\forall f, g \in \hat{\mathcal{Y}}$*, and* $\max_{\xi_1\in\mathbb{R}} \pi(\xi_1) = P < \infty$*. Let* $\varepsilon = \varepsilon_0/\sqrt{\gamma}$

with any $\varepsilon_0 > 0$. Then the minimizer u_γ of (3.3) converges to the solution u_* of the exact problem (2.1)–(2.3), and

$$\|u_\gamma - u_*\|_{\mathcal{U}}^2 \leq C\varepsilon_0^3\gamma^{-1/2} + \frac{1}{\alpha}\langle\lambda_*, Su_\gamma - y_{\max}\rangle_{\mathcal{Y}^*,\mathcal{Y}} \to 0, \qquad \gamma \to \infty,$$

where $C > 0$ is independent of γ and ε_0.

Remark 3.13. *For the classical Moreau-Yosida penalty with $\varepsilon_0 = 0$, we recover existing convergence estimates* [21, 2] *that depend only on $\langle\lambda_*, Su_\gamma - y_{\max}\rangle_{\mathcal{Y}^*,\mathcal{Y}}$. This term converges to 0 as shown in* (3.26)*, but the rate of this convergence can be estimated only if bounds on $\|\lambda_*\|_{L^2(\Omega,D)}$ or $\|Su_\gamma - y_{\max}\|_{\mathcal{Y}}$ can be established from other sources, such as the discretization of $\mathcal{Y}$* [21, Theorem 3.7].

Proof. We aim at refining the estimate (3.22). Specifically, we need to lower-bound $\langle(Su_\gamma - y_{\max})_-, g_\varepsilon(Su_\gamma - y_{\max})\rangle_{\mathcal{Y}^*,\mathcal{Y}}$, where $(y)_- = \min(y, 0)$. For brevity, let $f(x,\xi) = Su_\gamma - y_{\max}(x,\xi)$. Using the particular form of duality pairing and Assumption 3.1, we can write

$$\begin{aligned}\langle(Su_\gamma - y_{\max})_-, g_\varepsilon(Su_\gamma - y_{\max})\rangle_{\mathcal{Y}^*,\mathcal{Y}} &= \int_{\mathbb{R}^d}\int_D (f)_- g_\varepsilon(f)dx\pi(\xi_1)\cdots\pi(\xi_d)d\xi \\ &= \int_D \int_{f(x,\xi)\leq 0} f g_\varepsilon(f)\pi(\xi_1)\cdots\pi(\xi_d)d\xi dx. \qquad (3.27)\end{aligned}$$

Introduce a change of variables

$$\begin{bmatrix}\xi_1\\ \xi_2\\ \vdots\\ \xi_d\end{bmatrix} \to \begin{bmatrix}f(x,\xi)\\ \xi_2\\ \vdots\\ \xi_d\end{bmatrix}$$

with the Jacobian

$$J := \left|\det\begin{bmatrix}\frac{\partial f}{\partial\xi_1} & \frac{\partial f}{\partial\xi_2} & \cdots & \frac{\partial f}{\partial\xi_d}\\ 0 & 1 & \cdots & 0\\ & & \ddots & \\ 0 & \cdots & 0 & 1\end{bmatrix}\right| = \left|\frac{\partial f}{\partial\xi_1}\right| \geq c > 0.$$

Now we can express (3.27) using univariate integration,

$$\begin{aligned}\langle(Su_\gamma - y_{\max})_-, g_\varepsilon(Su_\gamma - y_{\max})\rangle_{\mathcal{Y}^*,\mathcal{Y}} &= \int_D\int_{\min f}^0\int_{\mathbb{R}^{d-1}} f g_\varepsilon(f)J^{-1}\pi(\xi_1(f))\cdots\pi(\xi_d)d\xi_2\cdots d\xi_d df dx \\ &\geq \int_D\int_{-\infty}^0 f g_\varepsilon(f)\frac{1}{c}P df dx \\ &\geq -|D|P\frac{1}{c}\varepsilon^3,\end{aligned}$$

where in the second line we used that the expression under the integral is nonpositive, and $\int\pi(x_2)dx_2 = \cdots = \int\pi(x_d)dx_d = 1$, and in the third line we used Lemma 3.11.

Now we can replace (3.22) as follows:

$$\langle \lambda_\gamma, Su_\gamma - Su_* \rangle_{\mathcal{Y}^*,\mathcal{Y}} \geq \gamma \left[0.5 \|(Su_\gamma - y_{\max})_+\|^2_{L^2(\Omega,D)} - 0.5|D|P\frac{1}{c}\varepsilon^3 \right].$$

Proceeding as in Theorem 3.10, we replace (3.25) by

$$\alpha\|u_\gamma - u_*\|^2_{\mathcal{U}} \leq \langle \lambda_*, Su_\gamma - y_{\max}\rangle_{\mathcal{Y}^*,\mathcal{Y}} + \frac{\gamma}{2}|D|P\frac{1}{c}\varepsilon^3.$$

Setting $\varepsilon = \varepsilon_0/\sqrt{\gamma}$, we obtain that

$$\|u_\gamma - u_*\|^2_{\mathcal{U}} \leq \frac{1}{\alpha}\langle \lambda_*, Su_\gamma - y_{\max}\rangle_{\mathcal{Y}^*,\mathcal{Y}} + \underbrace{\frac{|D|P}{2c\alpha}}_{C}\frac{\varepsilon_0^3}{\gamma^{1/2}}.$$

Thus the proof is complete. □

Remark 3.14. *This theorem can be generalized to vector-valued functions straightforwardly. Indeed, if $f_i(x,\xi)$ denotes the ith component of a vector function, the duality pairing* (3.27) *reads*

$$\langle (f)_-, g_\varepsilon(f)\rangle_{\mathcal{Y}^*,\mathcal{Y}} = \int_{\mathbb{R}^d}\int_D \sum_i (f_i)_- g_\varepsilon(f_i) dx\pi(\xi)d\xi = \sum_i \int_D \int_{f_i(x,\xi)\leq 0} f_i g_\varepsilon(f_i)\pi(\xi)d\xi dx,$$

and ξ_1 can be changed to f_i for each term of the sum over i.

The assumption of a lower bound of the Jacobian is practical. The Karhunen-Loeve expansion as in Example 3.2 is normally derived as the eigenvalue expansion of the covariance function of e.g. κ. By the Perron-Frobenius theorem, $\psi_1(x) = \partial\kappa/\partial\xi_1 > 0$. Further, $\partial y/\partial\kappa \neq 0$ due to ellipticity. Hence $\partial(Su)/\partial\xi_1 \neq 0$ whenever either u or boundary conditions or source term are nonzero. The remaining assumptions of Thm. 3.12 are also reasonable for practical solutions of regularized optimization problems. A convenient observation is that $\varepsilon = \varepsilon_0/\sqrt{\gamma}$ is the sufficient condition on the law of decay of the smoothing parameter for both Theorems 3.4 and 3.12.

4. Tensor-Train decomposition

Throughout this section, we use Assumption 3.1. Recall that the bottleneck is the computation of the expectation in e.g. gradient (3.5). While it may be possible to use a Monte Carlo quadrature, its convergence is usually slow, which may make estimates of small values of the gradient near the optimum particularly inaccurate. In this section, we describe the Tensor-Train (TT) decomposition as a function approximation technique that allows fast computation of the expectation. The original TT decomposition [30] was proposed for tensors (such as tensors of expansion coefficients), and the functional TT (FTT) decomposition [5, 19] has extended this idea to multivariate functions.

Let us introduce a basis $\{\ell_i(\xi_k)\}_{i=1}^{n_\xi}$ in each random variable ξ_k, $k = 1,\dots,d$, and a quadrature with nodes $Z = \{z_j\}$ and weights $\{w_j\}$ which is exact on this basis,

$$\mathbb{E}[\ell_i] = \sum_{j=1}^{n_\xi} w_j \ell_i(z_j).$$

For example, we can take Lagrange interpolation polynomials built upon a Gaussian quadrature, or orthogonal polynomials up to degree $n_\xi - 1$ together with the roots of the degree-n_ξ polynomial, or Fourier modes and the rectangular quadrature with the number of nodes corresponding to the highest frequency. Then we can approximate any random field $y \in \mathcal{Y}$ in the tensor product basis,

$$y(\xi) \approx \sum_{i_1=1}^{n_\xi} \cdots \sum_{i_d=1}^{n_\xi} \mathbf{Y}_{i_1,\ldots,i_d} \ell_{i_1}(\xi_1) \cdots \ell_{i_d}(\xi_d).$$

Note that the expansion coefficients $\mathbf{Y}$ form a tensor of n_ξ^d entries, which is impossible to store directly if d is large. The TT decomposition aims to factorize this tensor further to a product of tensors of manageable size.

Definition 4.1. *A tensor $\mathbf{Y} \in \mathbb{R}^{n_\xi \times \cdots \times n_\xi}$ is said to be approximated by the* TT decomposition *with a relative approximation error ϵ if there exist 3-dimensional tensors $\mathbf{Y}^{(k)} \in \mathbb{R}^{r_{k-1} \times n_\xi \times r_k}$, $k = 1, \ldots, d$, such that*

$$\tilde{\mathbf{Y}}_{i_1,\ldots,i_d} := \sum_{s_0,\ldots,s_d=1}^{r_0,\ldots,r_d} \mathbf{Y}^{(1)}_{s_0,i_1,s_1} \mathbf{Y}^{(2)}_{s_1,i_2,s_2} \cdots \mathbf{Y}^{(d)}_{s_{d-1},i_d,s_d}, \tag{4.1}$$

and $\|\mathbf{Y} - \tilde{\mathbf{Y}}\|_F = \epsilon \|\mathbf{Y}\|_F$. The factors $\mathbf{Y}^{(k)}$ are called TT cores*, and the ranges of summation indices $r_0, \ldots, r_d \in \mathbb{N}$ are called* TT ranks*. Note that without loss of generality we can let $r_0 = r_d = 1$.*

Plugging in the basis and redistributing the summations we obtain the FTT approximation

$$\tilde{y}(\xi) := \sum_{s_0,\ldots,s_d=1}^{r_0,\ldots,r_d} y^{(1)}_{s_0,s_1}(\xi_1) y^{(2)}_{s_1,s_2}(\xi_2) \cdots y^{(d)}_{s_{d-1},s_d}(\xi_d),$$

where

$$y^{(k)}_{s_{k-1},s_k}(\xi_k) = \sum_{i=1}^{n_\xi} \mathbf{Y}^{(k)}_{s_{k-1},i,s_k} \ell_i(\xi_k), \quad k = 1, \ldots, d.$$

Smooth [35], weakly correlated [33] or certainly structured [20] functions have been shown to induce rapidly converging TT approximations.

Given the TT decomposition, its expectation can be computed by first integrating each TT core, and then multiplying the TT cores one by one. Let

$$\mathbf{V}^{(k)}_{s_{k-1},s_k} = \sum_{j=1}^{n_\xi} w_j y^{(k)}_{s_{k-1},s_k}(z_j) = \sum_{i,j=1}^{n_\xi} w_j \mathbf{L}_{i,j} \mathbf{Y}^{(k)}_{s_{k-1},i,s_k}, \quad \text{where} \quad \mathbf{L}_{i,j} = \ell_i(z_j). \tag{4.2}$$

Now we multiply the matrices $\mathbf{V}^{(k)} \in \mathbb{R}^{r_{k-1} \times r_k}$ in order:

$$\mathbb{E}[\tilde{y}] = \left(\left(\left(\mathbf{V}^{(1)} \mathbf{V}^{(2)} \right) \mathbf{V}^{(3)} \right) \cdots \mathbf{V}^{(d)} \right). \tag{4.3}$$

Note that each step in (4.3) is a product of $1 \times r_{k-1}$ vector by $r_{k-1} \times r_k$ matrix. In turn, the univariate quadrature (4.2) requires $n_\xi^2 r_{k-1} r_k$ floating point operations if the Vandermonde matrix $\mathbf{L}$ is dense, and $n_\xi r_{k-1} r_k$ if it's sparse, for example, if Lagrange polynomials are used.

Introducing $r := \max_k r_k$, we conclude that the expectation of a TT decomposition can be computed with a complexity $\mathcal{O}(dr^2)$ which is linear in the dimension.

To compute a TT approximation, we employ the TT-Cross algorithm [31]. We start with an empirical risk minimization problem

$$\min_{\mathbf{Y}^{(1)},\dots,\mathbf{Y}^{(d)}} \sum_{j=1}^{N} \left(\tilde{y}(\xi^j) - y(\xi^j)\right)^2,$$

where $\Xi = \{\xi^j\}$ is a certain set of samples. To avoid minimization over all $\mathbf{Y}^{(1)}, \dots, \mathbf{Y}^{(d)}$ simultaneously (which is non-convex), we switch to an alternating direction approach: iterate over $k = 1, \dots, d$, solving in each step

$$\min_{\mathbf{Y}^{(k)}} \sum_{j=1}^{N} \left(\tilde{y}(\xi^j) - y(\xi^j)\right)^2. \tag{4.4}$$

This problem can be solved by linear normal equations. Indeed, introduce a matrix $\mathbf{Y}_{\neq k} \in \mathbb{R}^{N \times (r_{k-1} n_\xi r_k)}$ with elements

$$(\mathbf{Y}_{\neq k})_{j,t} = \sum_{s_0,\dots,s_{k-2}} y^{(1)}_{s_0,s_1}(\xi^j_1) \cdots y^{(k-1)}_{s_{k-2},s_{k-1}}(\xi^j_{k-1}) \ell_i(\xi^j_k) \sum_{s_{k+1},\dots,s_d} y^{(k+1)}_{s_k,s_{k+1}}(\xi^j_{k+1}) \cdots y^{(d)}_{s_{d-1},s_d}(\xi^j_d),$$

where $t = (s_{k-1} - 1) n_\xi r_k + (i-1) r_k + s_k$, and a vector $\mathbf{y}^{(k)} \in \mathbb{R}^{r_{k-1} n_\xi r_k}$ with elements $\mathbf{y}^{(k)}_t = \mathbf{Y}^{(k)}_{s_{k-1},i,s_k}$. Now $\tilde{y}(\Xi) = \mathbf{Y}_{\neq k} \mathbf{y}^{(k)}$, and (4.4) is minimized by

$$\mathbf{y}^{(k)} = (\mathbf{Y}^\top_{\neq k} \mathbf{Y}_{\neq k})^{-1} (\mathbf{Y}^\top_{\neq k} y(\Xi)). \tag{4.5}$$

To both select "good" sample set Ξ and simplify the assembly of $\mathbf{Y}_{\neq k}$, we restrict the set to have the Cartesian form

$$\Xi = \Xi_{<k} \times Z \times \Xi_{>k},$$

where $\Xi_{<k} = \{(\xi_1, \dots, \xi_{k-1})\}$, $\Xi_{>k} = \{(\xi_{k+1}, \dots, \xi_d)\}$ with *nestedness* conditions

$$(\xi_1, \dots, \xi_{k-1}, \xi_k) \in \Xi_{<k+1} \Rightarrow (\xi_1, \dots, \xi_{k-1}) \in \Xi_{<k},$$
$$(\xi_k, \xi_{k+1}, \dots, \xi_d) \in \Xi_{>k-1} \Rightarrow (\xi_{k+1}, \dots, \xi_d) \in \Xi_{>k}.$$

This makes

$$\mathbf{Y}_{\neq k} = \mathbf{Y}_{<k} \otimes \mathbf{L} \otimes \mathbf{Y}_{>k},$$

where

$$(\mathbf{Y}_{<k})_{j,s} = \sum_{s_0,\dots,s_{k-2}} y^{(1)}_{s_0,s_1}(\xi^j_1) \cdots y^{(k-1)}_{s_{k-2},s}(\xi^j_{k-1}), \qquad (\xi^j_1, \dots, \xi^j_{k-1}) \in \Xi_{<k},$$
$$(\mathbf{Y}_{>k})_{j,s} = \sum_{s_{k+1},\dots,s_d} y^{(k+1)}_{s,s_{k+1}}(\xi^j_{k+1}) \cdots y^{(d)}_{s_{d-1},s_d}(\xi^j_d), \qquad (\xi^j_{k+1}, \dots, \xi^j_d) \in \Xi_{>k}.$$

Moreover, $\mathbf{Y}_{<k+1}$ and $\mathbf{Y}_{>k-1}$ are submatrices of

$$\mathbf{Y}_{\leq k} := \begin{bmatrix} \mathbf{Y}_{<k} y^{(k)}(z_1) \\ \vdots \\ \mathbf{Y}_{<k} y^{(k)}(z_{n_\xi}) \end{bmatrix} \quad \text{and} \quad \mathbf{Y}_{\geq k} := \begin{bmatrix} y^{(k)}(z_1) \mathbf{Y}_{>k} & \cdots & y^{(k)}(z_{n_\xi}) \mathbf{Y}_{>k} \end{bmatrix}, \tag{4.6}$$

respectively. This allows us to build the sampling sets by selecting r_k rows of $\mathbf{Y}_{\leq k}$ (resp. columns of $\mathbf{Y}_{\geq k}$) by the *maximum volume principle* [18], which needs only $\mathcal{O}(n_\xi r^3)$ floating point operations per single matrix $\mathbf{Y}_{\leq k}$ or $\mathbf{Y}_{\geq k}$. The r_k indices of e.g. rows of $\mathbf{Y}_{\leq k}$ constituting the maximum volume submatrix $\mathbf{Y}_{<k}$ are also indices of the r_k tuples in $\Xi_{<k} \times Z$ constituting the next "left" set $\Xi_{<k+1}$. The "right" set $\Xi_{>k-1}$ is constructed analogously. This closes the recursion and allows us to carry out the alternating iteration in either direction, $k = 1, \ldots, d$ or $k = d, \ldots, 1$. By this construction, the cardinality of $\Xi_{<k+1}$ and $\Xi_{>k-1}$ is r_k. Hence, the cardinality of Ξ is $r_{k-1} n_\xi r_k$, and one full iteration of the TT-Cross algorithm needs $\mathcal{O}(d n_\xi r^2)$ samples of y.

One drawback of the "naive" TT-Cross algorithm outlines above is that the TT ranks are fixed. To adapt them to a desired error tolerance, several modifications have been proposed: merge ξ_k, ξ_{k+1} into one variable, optimize the corresponding larger TT core, and separate it into two actual TT cores using truncated singular value decomposition (SVD) [34] or matrix adaptive cross approximation [8]; oversample $\Xi_{<k}$ or $\Xi_{>k}$ with random or error-targeting points [10]; oversample the selection of submatrices from (4.6) by using the *rectangular* maximum volume principle [29].

However, in this paper we can pursue a somewhat more natural regression approach [7]. We will always need to approximate a vector function, where different components correspond to different degrees of freedom of an ODE or a PDE solution, or different components of a gradient. Since the procedure to evaluate y is now taking two arguments (ξ and, say, $m = 1, \ldots, M$ indexing extra degrees of freedom), we can replace the normal equations (4.5) by

$$\mathbf{y}^{(k)}(m) = (\mathbf{Y}_{\neq k}^\top \mathbf{Y}_{\neq k})^{-1} (\mathbf{Y}_{\neq k}^\top y(\Xi, m)),$$

which can be reshaped into a 4-dimensional tensor $\hat{\mathbf{Y}}^{(k)} \in \mathbb{R}^{r_{k-1} \times n_\xi \times r_k \times M}$ with elements $\hat{\mathbf{Y}}^{(k)}_{s_{k-1}, i, s_k, m} = \mathbf{y}^{(k)}_t(m)$. To compute the usual 3-dimensional TT core, we can use a simple Principal Component Analysis (PCA), which selects $\hat{r}$ slices $\mathbf{Y}^{(k)}_{s_{k-1}, i, 1}, \ldots, \mathbf{Y}^{(k)}_{s_{k-1}, i, \hat{r}}$ with the minimal $\hat{r}$ such that

$$\min_{\mathbf{W}} \sum_{s_{k-1}, i, s_k, m} \left(\sum_{s=1}^{\hat{r}} \mathbf{Y}^{(k)}_{s_{k-1}, i, s} \mathbf{W}_{s, s_k, m} - \hat{\mathbf{Y}}^{(k)}_{s_{k-1}, i, s_k, m} \right)^2 \leq \mathrm{tol}^2 \cdot \|\hat{\mathbf{Y}}^{(k)}\|_F^2.$$

Note that this problem is solved easily by the truncated SVD, where the new TT rank $\hat{r}$ can be chosen anywhere between 1 and $\min\{r_{k-1} n_\xi, r_k M\}$ to satisfy the error tolerance tol. After replacing r_k with $\hat{r}$, the TT-Cross iteration $k = 1, \ldots, d$ can proceed as previously. In the last step ($k = d$), the PCA step is omitted, and we obtain the so-called *block* TT decomposition [9], which in the functional form reads

$$\tilde{y}(\xi, m) = \sum_{s_0, \ldots, s_d} y^{(1)}_{s_0, s_1}(\xi_1) \cdots y^{(d-1)}_{s_{d-2}, s_{d-1}}(\xi_{d-1}) \hat{y}^{(d)}_{s_{d-1}, s_d}(\xi_d, m).$$

The "backward" iteration $k = d, \ldots, 1$ can be generalized similarly.

5. Practical computation of the smoothed Moreau-Yosida optimization

To compute the gradient of the cost function (3.5), we need to approximate the function under the expectation,

$$\mathbf{G}_u^{\varepsilon,h}(\xi) := \mathbf{S}_h(\xi)^* \cdot \mathrm{diag}(g'_\varepsilon(\mathbf{S}_h(\xi)u - \mathbf{y}^h_{\max}(\xi))) \cdot \mathbf{M} g_\varepsilon(\mathbf{S}_h(\xi)u - \mathbf{y}^h_{\max}(\xi)), \tag{5.1}$$

using the TT-Cross, followed by taking the expectation of the TT decomposition[1] This can be performed in two ways. To begin with, we can apply the TT-Cross algorithm to approximate directly $\mathbf{G}_u^{\varepsilon,h}(\xi)$. For each sample $\xi^j \in \Xi$, one needs to solve one forward problem to compute $\mathbf{S}_h(\xi^j)u$, and one adjoint problem to apply $\mathbf{S}_h(\xi^j)^*$ to the rest of the function. Recall that the TT-Cross needs $\mathcal{O}(dn_\xi r^2)$ samples, hence $\mathcal{O}(dn_\xi r^2)$ solutions of the forward, adjoint and sensitivity problems. However, the maximal TT rank r of the softplus and sigmoid functions typically grows proportional to $1/\varepsilon$. When the solution of the forward and adjoint problem is expensive (for example, in the PDE-constrained optimization), this may result in an excessive computational complexity.

Alternatively, we can first compute TT approximations $\tilde{\mathbf{y}}(\xi) \approx \mathbf{S}_h(\xi)u$ and $\tilde{\mathbf{S}}_h(\xi)^* \approx \mathbf{S}_h(\xi)^*$, followed by TT approximations $\tilde{\mathbf{g}}_\varepsilon(\xi) :\approx g_\varepsilon(\tilde{\mathbf{y}}(\xi) - \mathbf{y}^h_{\max}(\xi))$, $\tilde{\mathbf{g}}'_\varepsilon(\xi) :\approx g'_\varepsilon(\tilde{\mathbf{y}}(\xi) - \mathbf{y}^h_{\max}(\xi))$, and finally $\tilde{\mathbf{G}}_u^{\varepsilon,h}(\xi) \approx \tilde{\mathbf{S}}_h(\xi)^* \mathrm{diag}(\tilde{\mathbf{g}}'_\varepsilon(\xi))\tilde{\mathbf{g}}_\varepsilon(\xi)$ using the approximate solution $\tilde{\mathbf{y}}(\xi)$, which does not require the solution of the PDE anymore. The bottleneck now is the approximation of the matrix-valued function $\mathbf{S}_h(\xi)^* \in \mathbb{R}^{n_u \times n_y}$. If both n_y and n_u are large (for example, in a case of a distributed control), the computation of $\mathbf{S}_h(\xi)^*$ for each sample of ξ requires assembling this large dense matrix, equivalent to the solution of the adjoint problem with n_u right hand sides. Nevertheless, the tensor approximation of $\mathbf{S}_h(\xi)^*$ converges usually much faster (e.g. exponentially) compared to the approximation of $\mathbf{G}_u^{\varepsilon,h}(\xi)$ directly, hence the TT approximation of $\mathbf{S}_h(\xi)^*$ may need much smaller TT ranks compared to the TT approximation of $\mathbf{G}_u^{\varepsilon,h}(\xi)$. In turn, the TT-Cross applied to $\mathbf{S}_h(\xi)^*$ requires much fewer solutions of the forward problem. For a moderate n_u this makes it faster to precompute $\tilde{\mathbf{y}}(\xi)$ and $\tilde{\mathbf{S}}_h(\xi)^*$. The entire pseudocode of the smoothed Moreau-Yosida optimization is listed in Algorithm 1.

6. Numerical examples

We start with $\gamma_0 = 1$ and double $\gamma_{\ell+1} = 2\gamma_\ell$ in the course of the Newton iterations until a desired value of γ_* is reached. According to Theorem 3.4, we choose $\varepsilon_\ell = 0.5/\sqrt{\gamma_\ell}$. The iteration is stopped when γ_L has reached the maximal desired value γ_*, and the step size has become smaller than $\delta_{\min} = 10^{-3}$. We always take a zero control as the initial guess u_0, and $\theta = 10^{-4}$. All computations are carried out in MATLAB 2020b on a Intel Xeon E5-2640 v4 CPU, using TT-Toolbox (`https://github.com/oseledets/TT-Toolbox`).

6.1. One-dimensional Elliptic PDE.

We consider an elliptic PDE example from [22, 13]. Here, a misfit functional

$$j(u) = \frac{1}{2}\mathbb{E}\left[\|y(u,\omega,x) - y_d(x)\|^2_{L^2(D)}\right] + \frac{\alpha}{2}\|u(x)\|^2_{L^2(D)}$$

[1]Note that $\mathbf{G}_u^{\varepsilon,h}(\xi)$ is a vector function with M being the number of degrees of freedom in the discretized u.

Algorithm 1 Inexact projected Newton optimization with smoothed a.s. constraints

Require: Procedures to compute $\mathbf{S}_h u, j^h(u), \nabla_u j^h(u)$, constraint $\mathbf{y}^h_{\max}$, initial and maximal Moreau-Yosida parameters γ_0, γ_*, initial smoothing parameter ε_0, initial control u_0, approximation and stopping tolerance tol, maximal number of iterations L, Armijo tuning parameter $\theta \in (0,1)$, minimal step size $\delta_{\min} \in (0,1)$.

Ensure: Optimized control $u^{\gamma_*,h}$.

Set iteration number $\ell = 0$, step size $\delta = 1$, $u_{-1} = u_0$.
while $\ell < L$ **and** $\delta > \delta_{\min}$ **and** $\|u_\ell - u_{\ell-1}\|_{\mathcal{U}} > \text{tol} \cdot \|u_\ell\|_{\mathcal{U}}$ **or** $\ell = 0$ **or** $\gamma_\ell < \gamma_*$ **do**
  Set $\varepsilon = \varepsilon_0/\sqrt{\gamma_\ell}$.
  Approximate $\tilde{\mathbf{G}}^{\varepsilon,h}_{u_\ell}(\xi) \approx \mathbf{G}^{\varepsilon,h}_{u_\ell}(\xi)$ as shown in (5.1) using TT-Cross up to tolerance tol.
  Approximate $\tilde{\mathbf{g}}'_\varepsilon(\xi) \approx g'_\varepsilon(\mathbf{S}_h(\xi)u_\ell - \mathbf{y}^h_{\max}(\xi))$ using TT-Cross up to tolerance tol.
  Compute the gradient $\nabla_u j^{\gamma_\ell,\varepsilon,h} = \nabla_u j^h(u_\ell) + \gamma_\ell \mathbb{E}[\tilde{\mathbf{G}}^{\varepsilon,h}_{u_\ell}(\xi)]$
  Compute the anchor point $\xi_* = \mathbb{E}[\xi \cdot \mathbf{1}^\top \tilde{\mathbf{g}}'_\varepsilon(\xi)]/\mathbb{E}[\mathbf{1}^\top \tilde{\mathbf{g}}'_\varepsilon(\xi)]$.
  Compute the Newton direction $v = -\tilde{\mathbf{H}}^{-1}\nabla_u j^{\gamma_\ell,\varepsilon,h}$ using (3.9).
  Set step size $\delta = 1$.
  while $j^h(\mathcal{P}_{\mathcal{U}_{ad}}(u_\ell + \delta v)) > j^h(u_\ell) + \delta\theta\langle v, \nabla_u j^{\gamma_\ell,\varepsilon,h}\rangle_{\mathcal{U}^*,\mathcal{U}}$ **and** $\delta > \delta_{\min}$ **do**
    Set $\delta = \delta/2$.
  end while
  Set $u_{\ell+1} = \mathcal{P}_{\mathcal{U}_{ad}}(u_\ell + \delta v)$.
  Set $\gamma_{\ell+1} = \min\{2\gamma_\ell, \gamma_*\}$.
  Set $\ell = \ell + 1$.
end while
return $u^{\gamma_*,h} = u_\ell$.

is optimized subject to the stochastic PDE constraint[2]

$$\begin{aligned} \nu(\omega)\Delta y(u,\omega,x) &= g(\omega,x) + u(x), \quad (\omega,x) \in \Omega \times D, \\ \nu(\omega) &= 10^{\xi_1(\omega)-2}, \qquad g(\omega,x) = \frac{\xi_2(\omega)}{100}, \\ y|_{x=0} &= -1 - \frac{\xi_3(\omega)}{1000}, \qquad y|_{x=1} = -\frac{2+\xi_4(\omega)}{1000} \end{aligned} \tag{6.1}$$

where $D = (0,1)$, and $\xi(\omega) = (\xi_1(\omega), \ldots, \xi_4(\omega)) \sim \mathcal{U}(-1,1)^4$ is uniformly distributed. We take the desired state $y_d(x) = -\sin(50x/\pi)$ and the regularization parameter $\alpha = 10^{-2}$. Moreover, we add the constraints

$$y(u,\omega,x) \le y_{\max} \equiv 0 \quad \text{a.s.}, \quad \text{and} \quad -0.75 \le u(x) \le 0.75 \quad \text{a.e.}$$

We discretize (6.1) in the spatial coordinate x using linear finite elements on a uniform grid with n_y interior points, and in each random variable $\xi_i(\omega)$ using n_ξ Gauss-Legendre quadrature nodes on $(-1,1)$. Note that we exclude the boundary points $x = 0$ and $x = 1$ due to the Dirichlet boundary conditions. This spatial discretization is used for both y and u.

[2]Note that [22, 13] considered the constraint $y \ge 0$, so here we reverse the sign of y to make the constraint in the form (2.3).

Firstly, we study precomputation of the surrogate solution $\tilde{\mathbf{y}}(\xi)$ and adjoint operator $\tilde{\mathbf{S}}_h^*(\xi)$. We fix $n_y = 63$, $n_\xi = 65$, the TT approximation tolerance 10^{-7} and the final Moreau-Yosida regularization parameter $\gamma_* = 1000$. The direct computation of the TT approximation of (5.1) requires 995 seconds of the CPU time due to the maximal TT rank of 87. In contrast, $\tilde{\mathbf{S}}_h^*$ has the maximal TT rank of 8, and the computation of $\tilde{\mathbf{S}}_h^*$ requires only 64 seconds despite a larger $n_y \times n_y$ TT core carrying the spatial variables. Using the surrogates $\tilde{\mathbf{y}}$ and $\tilde{\mathbf{S}}_h^*$, the remaining computation of $\nabla_u j^{\gamma,\varepsilon,h}$ can be completed in less than 15 seconds. The relative difference between the two approximations of $\nabla_u j^{\gamma,\varepsilon,h}$ is below the TT approximation tolerance. This shows that the surrogate forward solution can significantly speed up Algorithm 1 without degrading its convergence, so we use it in all remaining experiments in this subsection.

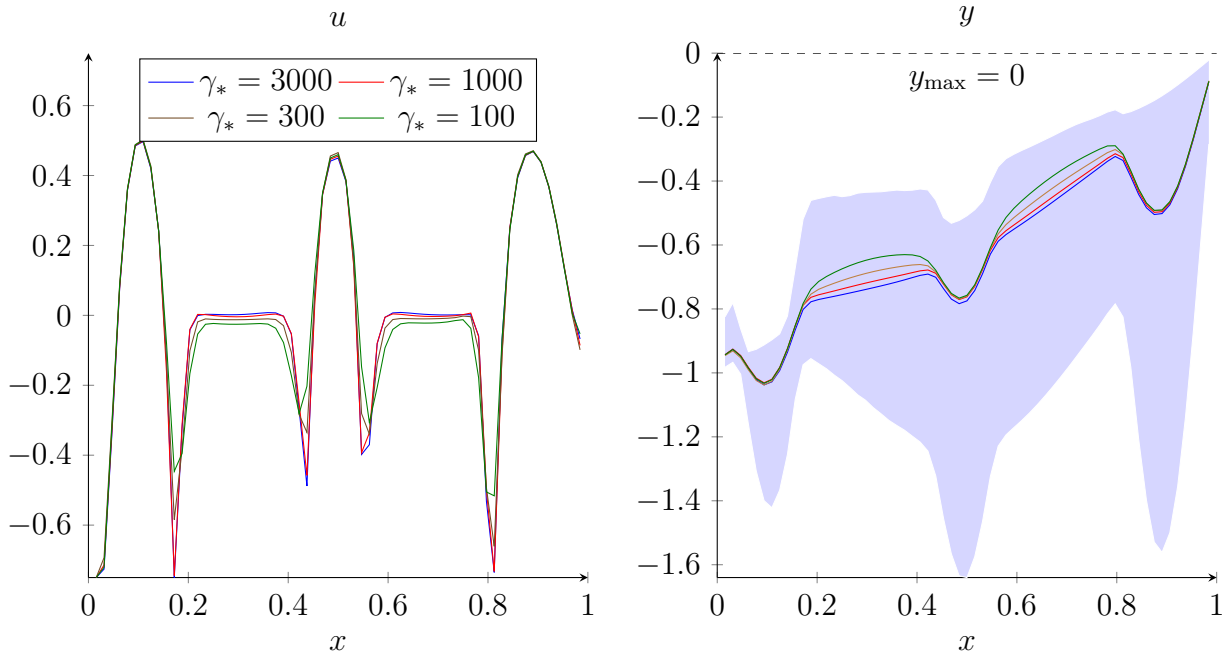

FIGURE 1. Left: control signals $u(x)$ for different γ_*. Right: mean (solid lines) and 95% confidence interval (shaded area, for $\gamma_* = 3000$ only) of the state $y(u^{\gamma_*}, \omega, x)$.

In Figure 1 we show the solutions (control and state) for varying final Moreau-Yosida penalty parameter γ_*, fixing $n_y = 63$, $n_\xi = 129$ and the TT approximation tolerance of 10^{-6}. We see that the solution converges with increasing γ_*, and larger γ_* yields a smaller probability of the constraint violation, albeit at a larger misfit cost $j(u)$, as shown in Figure 2. In particular, $\gamma_* > 300$ gives a solution with less than 1% of the constraint violation, such that the empirical 95% confidence interval computed using 1000 samples of the converged field $y(u^{\gamma_*})$ (see Fig. 1, right) is entirely within the constraint.

Finally, we study the convergence in the approximation parameters more systematically in Figure 3. In each plot we fix two out of three parameters: the final Moreau-Yosida penalty γ_*, the number of discretization points in the random variables n_ξ, and the number

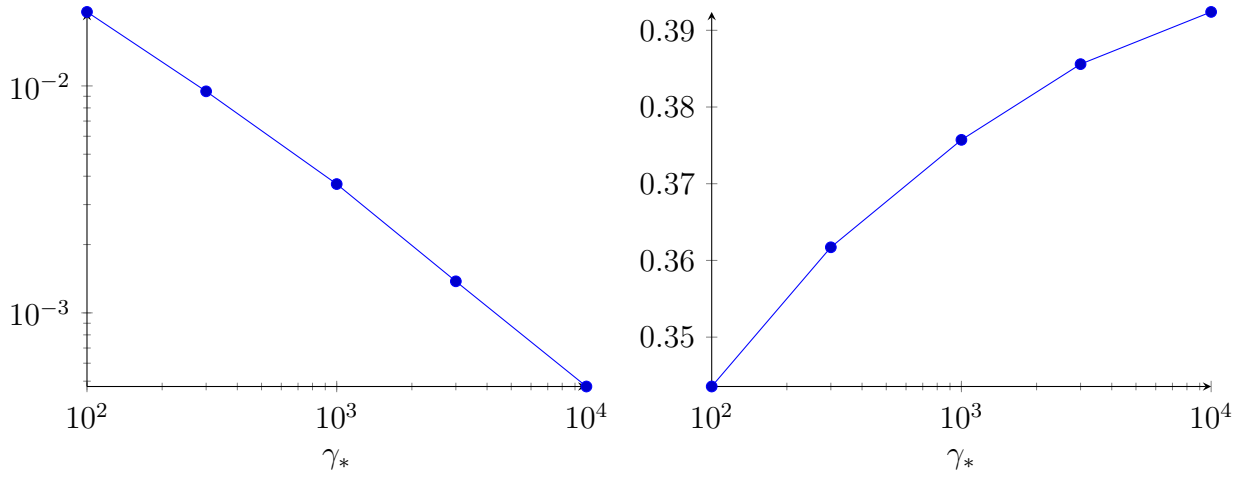

FIGURE 2. Left: probability of the constraint violation, $\mathbb{P}(y(u^{\gamma_*}, \omega, x) > 0)$. Right: total final cost $j(u^{\gamma^*})$.

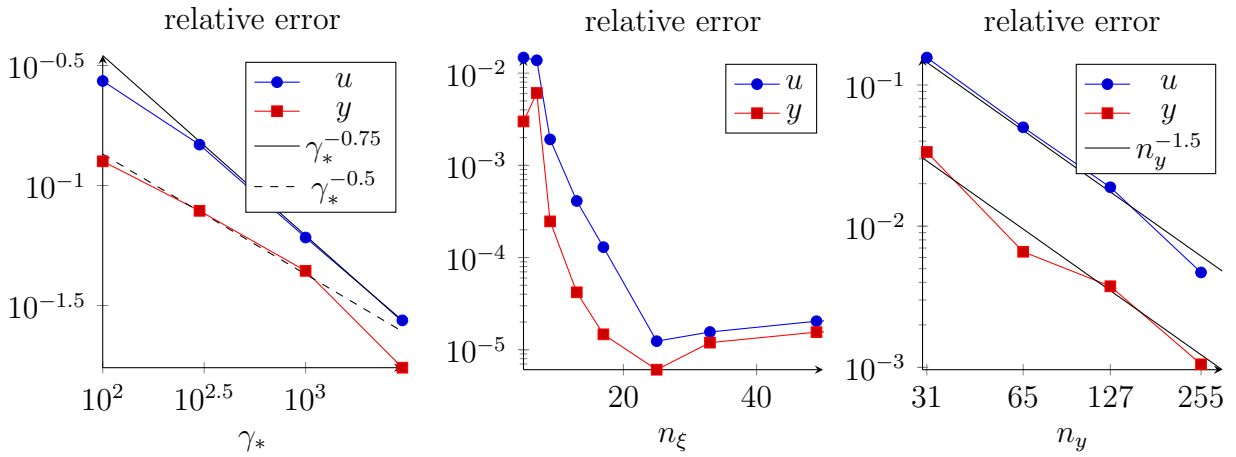

FIGURE 3. Relative L^2-norm difference from y and u to the reference solutions with $\gamma_* = 10^4$ with fixed $n_\xi = 257$, $n_y = 63$ (left), $n_\xi = 129$ with fixed $\gamma_* = 100$, $n_y = 63$ (middle) and $n_y = 511$ with fixed $\gamma_* = 100$, $n_\xi = 25$ (right).

of discretization points in space n_y. In addition, we fix the TT approximation threshold to 10^{-8} to reduce its influence. We observe a convergence in line with the $\gamma_*^{-1/2}$ rate of Theorem 3.4, exponential in n_ξ (which is often the case for a polynomial approximation of smooth functions [37]) until the tensor approximation error is hit, and between first and second order in n_y, which seems to be an interplay of the discretization consistency of the linear finite elements (second order) and box constraints (first order).

6.2. **Two-dimensional elliptic PDE.** Now consider a two-dimensional extension of the previous problem,

$$\nu(\omega)\Delta y(u,\omega,x) = g(\omega,x) + u(x), \qquad (\omega,x) \in \Omega \times D, \tag{6.2}$$

$$y|_{x_1=0} = b_1(\omega)(1-x_2) + b_2(\omega)x_2, \qquad y|_{x_2=1} = b_2(\omega)(1-x_1) + b_3(\omega)x_1 \tag{6.3}$$

$$y|_{x_1=1} = b_4(\omega)(1-x_2) + b_3(\omega)x_2, \qquad y|_{x_2=0} = b_1(\omega)(1-x_1) + b_4(\omega)x_1, \tag{6.4}$$

$$\nu(\omega) = 10^{\xi_1(\omega)-2}, \qquad g(\omega,x) = \frac{\xi_2(\omega)}{100}, \tag{6.5}$$

$$b_1(\omega) = -1 - \frac{\xi_3(\omega)}{1000}, \qquad b_2(\omega) = -\frac{2+\xi_4(\omega)}{1000}, \tag{6.6}$$

$$b_3(\omega) = -1 - \frac{\xi_5(\omega)}{1000}, \qquad b_4(\omega) = -\frac{2+\xi_6(\omega)}{1000}, \tag{6.7}$$

where $D = (0,1)^2$, and $\xi(\omega) = (\xi_1(\omega), \ldots, \xi_6(\omega)) \sim \mathcal{U}(-1,1)^6$ is uniformly distributed. We optimize the regularized misfit functional

$$j(u) = \frac{1}{2}\mathbb{E}\left[\|y(u,\omega,x) - y_d(x)\|^2_{L^2(D)}\right] + \frac{\alpha}{2}\|u(x)\|^2_{L^2(D)}$$

with the desired state $y_d(x) = -\sin(50x_1/\pi)\cos(50x_2/\pi)$ and the regularization parameter $\alpha = 10^{-2}$, subject to constraints

$$y(u,\omega,x) \le y_{\max} \equiv 0 \quad \text{a.s.}, \quad \text{and} \quad -0.75 \le u(x) \le 0.75 \quad \text{a.e.}$$

We smooth the almost sure constraint by the Moreau-Yosida method with the ultimate penalty parameter $\gamma_* = 10^2$.

We discretize both y and u in (6.2) using bilinear finite elements on a $n_y \times n_y$ rectangular grid. For the two-dimensional problem, the operator $\tilde{\mathbf{S}}_h^*$ is a dense matrix of size $n_y^2 \times n_y^2$, which we are unable to precompute. Therefore, we use the TT-Cross to approximate $\mathbf{G}_u^{\varepsilon,h}(\xi)$ directly.

In Figure 4 we show the optimal control, mean and standard deviation of the solution for $n_y = 63$ and $n_\xi = 17$. We see that the mean solution reflects the desired state subject to the constraints. The final cost $j(u^{\gamma_*})$ is about 0.222634, and the probability of the constraint violation is 0.0139223. The Newton method took $L = 37$ iterations to converge, the maximal TT rank of $\tilde{\mathbf{y}}(\xi)$ was 10 which was the same in all iterations, the maximal rank of $g'_\varepsilon(\tilde{\mathbf{y}} - \mathbf{y}^h_{\max})$ was 300, attained at the iteration after reaching γ_* (iteration 9), and the maximal rank of $\tilde{\mathbf{G}}_u^{\varepsilon,h}(\xi)$ was 56 (in the final iterations). The computation took about a day of CPU time. However, these TT ranks are comparable to those in the one-dimensional example. This shows that the proposed technique can be also applied to a high-dimensional physical space, including complex domains and non-uniform grids, since the TT structure is independent of the spatial discretization.

6.3. **Variational inequality constraints.** In this section we minimize the regularized misfit

$$j(u) = \frac{1}{2}\mathbb{E}[\|y(u,\omega,x) - y_d(x)\|^2_{L^2(D)}] + \frac{1}{2}\|u(x)\|^2_{L^2(D)} \tag{6.8}$$

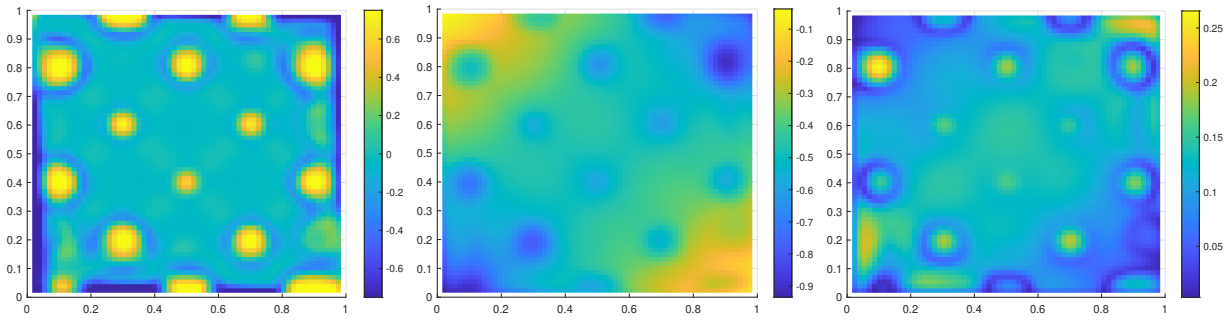

FIGURE 4. Left: control signal $u^{\gamma*}(x)$. Middle: mean $\mathbb{E}[y(u^{\gamma*},\omega,x)]$. Right: standard deviation $\sqrt{\mathbb{E}[(y(u^{\gamma*},\omega,x)-\mathbb{E}[y(u^{\gamma*},\omega,x)])^2]}$.

subject to a random elliptic variational inequality (VI) constraint,

$$y(u,\omega,x)\leq 0: \quad \langle A(\omega)y(u,\omega,x)-f(\omega,x)-B(\omega,x)u, y(u,\omega,x)-v\rangle \leq 0, \quad \forall v: v\leq 0. \tag{6.9}$$

We use Example 5.1 from [1] (with the reversed sign of y), where $D=(0,1)^2$, $A=-\Delta$, $B=\mathrm{Id}$, and deterministic functions constructing the desired state:

$$\hat{y}(x)=\begin{cases}160(x_1^3-x_1^2+0.25x_1)(x_2^3-x_2^2+0.25x_2) & \text{in } (0,0.5)^2,\\ 0, & \text{otherwise,}\end{cases}$$
$$\hat{\zeta}(x)=\max(0,-2|x_1-0.8|-2|x_1x_2-0.3|+0.5),$$
$$y_d(x)=-\hat{y}-\hat{\zeta}+\Delta\hat{y}.$$

In contrast, the right hand side depends on the random variables,

$$f(\xi(\omega),x)=\Delta\hat{y}+\hat{y}+\hat{\zeta}+b(\xi(\omega),x),$$
$$b(\xi(\omega),x)=\begin{cases}\sum_{i=1}^{d}\sqrt{\lambda_i}\phi_i(x)\xi_i(\omega), & \text{in } (0,0.5)\times(0,1),\\ 0, & \text{otherwise.}\end{cases}$$

The Karhunen-Loeve expansion in $b(\xi,x)$ is an affine-uniform random field, with $\xi_i(\omega)\sim\mathcal{U}(-1,1)$, $\phi_i(x)=2\cos(\pi jx_2)\cos(\pi kx_1)$ and $\lambda_i=\frac{1}{100}\exp(-\frac{\pi}{4}(j^2+k^2))$, where the pairs (j,k), $j,k=1,2,\ldots$, are permuted such that $\lambda_1\geq\lambda_2\geq\cdots$.

The VI (6.9) is replaced by the penalized problem

$$Ay+\frac{1}{\varepsilon}g_\varepsilon(y)=f(\xi,x)+Bu, \tag{6.10}$$

so we minimize (6.8) with $y(u,\xi,x)$ plugged in from (6.10). The latter equation is solved via the Newton method, initialized with $y=0$ as the initial guess, and stopped when the relative difference between two consecutive iterations of y falls below 10^{-12}. The problem is discretized in x via the piecewise bilinear finite elements on a uniform $n_y\times n_y$ grid with cell size $h=1/(n_y+1)$. The homogeneous Dirichlet boundary conditions $y=0$ on ∂D allow us to store only interior grid points. This gives us a discrete problem of minimizing

$$j^h(\mathbf{u})=\frac{1}{2}\mathbb{E}[\|\mathbf{y}(\mathbf{u},\xi)-\mathbf{y}_d\|^2_{\mathbf{M}_h}]+\frac{1}{2}\|\mathbf{u}\|^2_{\mathbf{M}_h} \tag{6.11}$$

subject to

$$\mathbf{A}_h\mathbf{y} + \frac{1}{\varepsilon}g_\varepsilon(\mathbf{y}) = \mathbf{f}(\xi) + \mathbf{u}, \tag{6.12}$$

where $\mathbf{A}_h, \mathbf{M}_h \in \mathbb{R}^{n_y^2 \times n_y^2}$ are the stiffness and mass matrices, respectively.

The state part of the cost

$$j_y(\mathbf{u}, \xi) = \frac{1}{2}\|\mathbf{y}(\mathbf{u}, \xi) - \mathbf{y}_d\|^2_{\mathbf{M}_h}$$

and its gradient

$$\nabla_u j_y(\mathbf{u}, \xi) = \mathbf{S}_h^*(\xi)\mathbf{M}_h(\mathbf{y}(\mathbf{u}, \xi) - \mathbf{y}_d)$$

are approximated by the TT-Cross (as functions of ξ), which allows one to compute the expectation of $\tilde{j}_y(\mathbf{u}, \xi) \approx j_y(\mathbf{u}, \xi)$ and $\nabla_u \tilde{j}_y(\mathbf{u}, \xi) \approx \nabla_u j_y(\mathbf{u}, \xi)$ easily. The forward model (6.12) is solved at each evaluation of ξ in the TT-Cross. However, to avoid excessive computations, the Hessian of (6.11) is approximated by that anchored at the mean point $\xi = 0$:

$$\nabla_{uu} j^h(\mathbf{u}) \approx \tilde{\mathbf{H}} := \mathbf{S}_h^*(0)\mathbf{M}_h\mathbf{S}_h'(0) + \mathbf{M}_h.$$

The Newton system $\tilde{\mathbf{H}}^{-1}\nabla_u j^h$ is solved iteratively by using the CG method, since the matrix-vector product with $\tilde{\mathbf{H}}$ requires the solution of only one forward and one adjoint problem,

$$\mathbf{S}_h^* \cdot \mathbf{v} = \mathbf{S}_h' \cdot \mathbf{v} = \left(\mathbf{A}_h + \operatorname{diag}\left(\frac{1}{\varepsilon}g_\varepsilon'(\mathbf{y})\right)\right)^{-1}\mathbf{v}, \quad \forall \mathbf{v} \in \mathbb{R}^{n_y^2}. \tag{6.13}$$

In Table 1 we vary the dimension of the random variable d, the number of quadrature points in each random variable n_ξ, and the approximation tolerance in the TT-Cross (tol). The spatial grid size is fixed to $n_y = 31$, which is comparable with the resolution in [1], and the smoothing parameter $\varepsilon = 10^{-6}$. As a reference solution $\mathbf{u}_*$, we take the control computed with $d = 20$, $n_\xi = 5$ and $\text{tol} = 10^{-4}$. We see that the control and the cost can be approximated quite accurately even with a very low order of the polynomial approximation in ξ. It also seems unnecessary to keep 20 terms in the Karhunen-Loeve expansion.

The computation complexity is dominated by the solutions of the forward and adjoint problems. The article [1] reports a "# PDE solves" in a path-following stochastic variance reduced gradient method solving (6.8)–(6.9). We believe this indicates the number of the complete solutions of the PDE (6.12). However, each solution of (6.12) to the increment tolerance 10^{-12} requires 23–25 Newton iterations, each of which requires the linear system solution of the form (6.13), Moreover, the anchored outer Hessian $\tilde{\mathbf{H}}$ requires two extra linear solves. Therefore, in Table 1, we show both the number of PDE solutions till convergence, N_{pde}, and the number of all linear system solutions N_{lin}, occurred during the optimization of (6.11) till the relative increment of $\mathbf{u}$ falls below the TT-Cross tolerance. In addition, we report the maximal TT ranks of the state cost gradient and the state itself. Note that assembly of the full state is not needed during the optimization of (6.11) – only certain samples of $\mathbf{y}(\mathbf{u}, \xi)$ are needed in the TT-Cross approximation of $\nabla_u j^h$. To save the computing time, the TT tensor of the entire state is computed only after the optimization of $\mathbf{u}$ has converged.

In Figure 5 we show the mean optimized forward state and the control. The results coincide qualitatively with those in [1]. If we consider the computational cost necessary to

TABLE 1. Cost, error in the control, number of solutions of $n_y^2 \times n_y^2$ linear system as in (6.13), number of complete forward PDE solutions (6.12), and the TT ranks of the cost gradient and forward solution.

d	n_ξ	tol	$j^h(\mathbf{u})$	$\frac{\|\mathbf{u}-\mathbf{u}_*\|_{\mathbf{M}_h}}{\|\mathbf{u}_*\|_{\mathbf{M}_h}}$	N_{lin}	N_{pde}	$r(\nabla_u \tilde{j}_y)$	$r(\tilde{\mathbf{y}})$
10	5	10^{-4}	1.261333069	1.1473e-06	1070007	44584	85	316
20	3	10^{-3}	1.261333069	2.9012e-05	46312	1976	7	29
20	3	10^{-4}	1.261333069	4.2713e-06	433134	18153	56	183
20	5	10^{-4}	1.261333069	—	1840467	76243	102	402

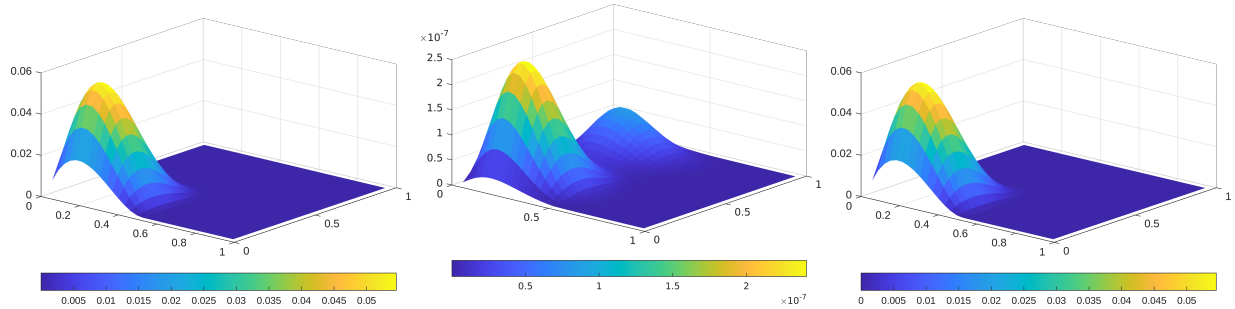

FIGURE 5. Left: mean optimised state $\mathbb{E}[-y]$ with $d = 20$, $n_\xi = 3$ and tol $= 10^{-3}$. Middle: variance of the optimized state $\mathbb{E}[(y - \mathbb{E}[y])^2]$. Right: optimised control u.

compute the optimal control only, we can notice that N_{pde} is significantly lower than the 291808 PDE solves in the stochastic variance reduced gradient method of [1].

6.4. **SEIR ODE model.** Now consider a slightly simplified version of the epidemiological ODE model used for the propagation of COVID-19 in the UK using the data from March-May 2020 [11]. This is a compartmental differential equation model with the following compartments.

- Susceptible (S).
- Exposed (E), but not yet infectious.
- Infected SubClinical type 1 (I^{SC1}): may require hospitalization in the future.
- Infected SubClinical type 2 (I^{SC2}): will recover without hospitalization.
- Infected Clinical type 1 (I^{C1}): individuals in the hospital who may decease.
- Infected Clinical type 2 (I^{C2}): individuals in the hospital who will recover.
- Recovered (R) and immune to reinfections.
- Deceased (D).

In turn, each of these compartments are split into 5 further sub-compartments corresponding to age bands: 0-19, 20-39, 40-59, 60-79 and 80+. The number of individuals in each compartment is denoted by the name of the compartment and age band index, For example, S_i denotes the number of susceptible individuals in the ith age band ($i = 1, \ldots, 5$), E_i denotes the number of exposed individuals in the ith age band, and so on. Variables corresponding

to different age bands but same compartment are collected into vectors, $S = (S_1, \dots, S_5)$, $E = (E_1, \dots, E_5)$ and so on.

Some of the variables introduced above are coupled to others only one way, and can be removed from the actual simulations. First, when the number of infected individuals is small compared to the population size (which is typically the case in the early stages of the epidemic), the relative variation of S is small. Hence, S can be taken constant instead of solving an ODE on it. Similarly, none of the variables depend on R and D, so they can be excluded from a coupled system of ODEs too, and computed separately after the solution of the ODEs. With these considerations in mind, the forward model reads as follows:

$$\frac{d}{dt}\begin{bmatrix} E \\ I^{SC1} \\ I^{SC2} \\ I^{C1} \\ I^{C2} \end{bmatrix} - \begin{bmatrix} -\kappa \mathtt{I} & A_u & A_u & 0 & 0 \\ \kappa \cdot \mathrm{diag}(\rho) & -\eta_C \mathtt{I} & 0 & 0 & 0 \\ \kappa \cdot \mathrm{diag}(1-\rho) & 0 & -\eta_R \mathtt{I} & 0 & 0 \\ 0 & \eta_C \cdot \mathrm{diag}(\rho') & 0 & -\nu \mathtt{I} & 0 \\ 0 & \eta_C \cdot \mathrm{diag}(1-\rho') & 0 & 0 & -\eta_{R,C} \mathtt{I} \end{bmatrix} \begin{bmatrix} E \\ I^{SC1} \\ I^{SC2} \\ I^{C1} \\ I^{C2} \end{bmatrix} = 0. \tag{6.14}$$

Here $\mathtt{I} \in \mathbb{R}^{5\times 5}$ is the identity matrix and $\mathrm{diag}(\cdot)$ produces a diagonal matrix from a vector. The control is defined in terms of the intensity of lockdown measures, and affects the susceptible-infected interaction matrix $A_u = \chi \cdot \mathrm{diag}(S) \cdot C_u \cdot \mathrm{diag}(\frac{1}{N})$, where

$$C_u = \mathrm{diag}(c^{home}) C^{home} + \mathrm{diag}(c_u^{work}) C^{work} + \mathrm{diag}(c_u^{school}) C^{school} + \mathrm{diag}(c_u^{other}) C^{other} \tag{6.15}$$

is the matrix of contact intensities between the age compartments. The total contact intensity is a sum of pre-pandemic contact intensity matrices in the four setting $C^{home}, C^{work}, C^{school}$ and C^{other}, multiplied by the reduction factors $c^{home}, c_u^{work}, c_u^{school}$ and c_u^{other} due to the lockdown measures. Since home contacts cannot be controlled, $c^{home} = (1, \dots, 1)$, but the remaining factors vary proportionally to the lockdown control applied from day 17 onwards,

$$c_u^{\mu}(t) = \begin{cases} (1,1,1,1,1)^\top, & t < 17, \\ (c_{123}(1-u^\mu(t)), c_{123}(1-u^\mu(t)), c_{123}(1-u^\mu(t)), c_4, c_5)^\top, & 17 \le t \le 90, \\ (c_{123}(1-u^\mu(90)), c_{123}(1-u^\mu(90)), c_{123}(1-u^\mu(90)), c_4, c_5)^\top, & t > 90, \end{cases} \tag{6.16}$$

where $\mu \in \{work, school, other\}$, u^μ are the intensities of lockdown measures applied to each setting μ, and c_{123}, c_4, c_5 are the initial contact intensities in the corresponding age groups. Note that the control will be optimized only on the time interval $[17, 90]$. Before day 17 the contact intensities are not reduced (no lockdown). From day 90 onwards we continue applying the last value of the control.

In addition, the model depends on the following parameters:

- χ: probability of S–I^{SC} interactions.
- $\kappa = 1/d_L$: average rate of an Exposed individual becoming SubClinical. It is inversely proportional to the average number of days d_L an individual stays in the Exposed state.
- $\eta_C = 1/d_C$: average rate of a SubClinical individual becoming Clinical. Similarly, d_C is the average time spent in the SubClinical state.
- $\eta_R = 1/d_R$: rate of recovery from I^{SC2}.
- $\eta_{R,C} = 1/d_{R,C}$: rate of recovery from I^{C2}.

- $\nu = 1/d_D$: rate of decease in the I^{C1} state.
- $\rho = (\rho_1, \ldots, \rho_5)^\top \in \mathbb{R}^5$: correction coefficients of the Exposed $\to$ SubClinical 1 transition rate for different age bands.
- $\rho' = (\rho'_1, \ldots, \rho'_5)^\top \in \mathbb{R}^5$: correction coefficients of the SubClinical $\to$ Clinical 1 transition.
- $N = (N_1, \ldots, N_5)^\top \in \mathbb{R}^5$: total number of individuals in each age group.
- N^0: total number of infected individuals on day 0.
- $N^{in} = (0.1, 0.4, 0.35, 0.1, 0.05)^\top N^0$: age partition of the initial number of infected individuals.

The ODE (6.14) is initialized by setting

$$E(0) = \frac{N^{in}}{3}, \quad I^{SC1}(0) = \frac{2}{3}\mathrm{diag}(\rho)N^{in}, \quad I^{SC2}(0) = \frac{2}{3}\mathrm{diag}(1-\rho)N^{in}, \quad I^{C1}(0) = I^{C2}(0) = 0.$$

The population sizes $S = N$ are taken from the Office for National Statistics, mid 2018 estimate.

However, none of the model parameters above are known beforehand. In [11], those were treated as random variables, and their distributions were estimated from observed numbers of infections and hospitalizations during the first 90 days using Approximate Bayesian Computation (ABC). In general, these variables are correlated through the posterior distribution, sampling from which is a daunting problem. Here, we replace the joint ABC posterior distribution by independent uniform distributions with a scaled posterior standard deviation centered around the posterior mean:

$$\begin{aligned}
\chi &\sim \mathcal{U}(0.13 - 0.03\sigma, 0.13 + 0.03\sigma), & d_L &\sim \mathcal{U}(1.57 - 0.42\sigma, 1.57 + 0.42\sigma), \\
d_C &\sim \mathcal{U}(2.12 - 0.80\sigma, 2.12 + 0.80\sigma), & d_R &\sim \mathcal{U}(1.54 - 0.40\sigma, 1.54 + 0.40\sigma), \\
d_{R,C} &\sim \mathcal{U}(12.08 - 1.51\sigma, 12.08 + 1.51\sigma), & d_D &\sim \mathcal{U}(5.54 - 2.19\sigma, 5.54 + 2.19\sigma), \\
\rho_1 &\sim \mathcal{U}(0.06 - 0.03\sigma, 0.06 + 0.03\sigma), & \rho_2 &\sim \mathcal{U}(0.05 - 0.03\sigma, 0.05 + 0.03\sigma), \\
\rho_3 &\sim \mathcal{U}(0.08 - 0.04\sigma, 0.08 + 0.04\sigma), & \rho_4 &\sim \mathcal{U}(0.54 - 0.22\sigma, 0.54 + 0.22\sigma), \\
\rho_5 &\sim \mathcal{U}(0.79 - 0.14\sigma, 0.79 + 0.14\sigma), & \rho'_1 &\sim \mathcal{U}(0.26 - 0.23\sigma, 0.26 + 0.23\sigma), \\
\rho'_2 &\sim \mathcal{U}(0.28 - 0.25\sigma, 0.28 + 0.25\sigma), & \rho'_3 &\sim \mathcal{U}(0.33 - 0.27\sigma, 0.33 + 0.27\sigma), \\
\rho'_4 &\sim \mathcal{U}(0.26 - 0.11\sigma, 0.26 + 0.11\sigma), & \rho'_5 &\sim \mathcal{U}(0.80 - 0.13\sigma, 0.80 + 0.13\sigma), \\
N^0 &\sim \mathcal{U}(276 - 133\sigma, 276 + 133\sigma), & c_{123} &\sim \mathcal{U}(0.63 - 0.21\sigma, 0.63 + 0.21\sigma), \\
c_4 &\sim \mathcal{U}(0.57 - 0.23\sigma, 0.57 + 0.23\sigma), & c_5 &\sim \mathcal{U}(0.71 - 0.23\sigma, 0.71 + 0.23\sigma).
\end{aligned} \tag{6.17}$$

Here, σ is the standard deviation scaling parameter, taken to be 0.03 in our experiment. This distribution behaves qualitatively similar to the posterior distribution in the vicinity of the posterior mean. It provides sufficient randomness to benchmark the constrained optimization method, while admitting independent sampling and gridding, needed for the TT approximations. That is, (6.17) form a random vector

$$\xi = (\chi, d_L, d_C, d_R, d_{R,C}, d_D, \rho_1, \rho_2, \rho_3, \rho_4, \rho_5, \rho'_1, \rho'_2, \rho'_3, \rho'_4, \rho'_5, N^0, c_{123}, c_4, c_5)$$

of $d = 20$ independent random variables, the state vector is

$$y(\xi, t) = (E_1, \ldots, E_5,\ I_1^{SC1}, \ldots, I_5^{SC1},\ I_1^{SC2}, \ldots, I_5^{SC2},\ I_1^{C1}, \ldots, I_5^{C1},\ I_1^{C2}, \ldots, I_5^{C2}),$$

and the ODE (6.14) constitutes the forward problem.

For the inverse problem, we use the total number of deceased patients as the cost function. The rate of decease is proportional to the number of Clinical type 1 individuals, so the total number of deceased individuals can be computed as

$$D(\xi, t) = \nu \int_0^t I^{C1}(\xi, s) ds. \tag{6.18}$$

To regularize the problem, we add also the norm of the control $u(t) = (u^{work}(t), u^{school}(t), u^{other}(t))$. Thus, the total cost function reads

$$j(u) = \frac{1}{2}\mathbb{E}[D(\xi, T)] + \frac{\alpha}{2} \int_{17}^{90} \|u(t)\|_2^2 dt, \tag{6.19}$$

where $T = 100$ is the final simulation time, and α is the regularization parameter, which we set to 100 in our experiment. Note that the norm of the control is taken only over the time interval $[17, 90]$ where the control varies.

We introduce the following constraints. Firstly, we limit the control components to the intervals $u^{work} \in [0, 0.69]$, $u^{school} \in [0, 0.9]$ and $u^{other} \in [0, 0.59]$. Next, we constrain the $\mathcal{R}$ number at the end of the variable control interval, $\mathcal{R}(\xi, 90) \leq 1$. In our model, the $\mathcal{R}$ number can be computed as $\mathcal{R}(\xi, t) = \lambda_{\max}(K)$, where

$$K = - \begin{bmatrix} 0 & A_u & A_u & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -\kappa \mathtt{I} & 0 & 0 & 0 & 0 \\ \kappa \cdot \mathrm{diag}(\rho) & -\eta_C \mathtt{I} & 0 & 0 & 0 \\ \kappa \cdot \mathrm{diag}(1-\rho) & 0 & -\eta_R \mathtt{I} & 0 & 0 \\ 0 & \eta_C \cdot \mathrm{diag}(\rho') & 0 & -\nu \mathtt{I} & 0 \\ 0 & \eta_C \cdot \mathrm{diag}(1-\rho') & 0 & 0 & -\eta_{R,C} \mathtt{I} \end{bmatrix}^{-1},$$

and $\lambda_{\max}$ denotes the maximal in modulus eigenvalue. Recall that $\mathcal{R} < 1$ implies that the epidemic decays, while $\mathcal{R} > 1$ corresponds to an expanding epidemic. The full smoothed Moreau-Yosida cost function becomes

$$j^{\gamma,\varepsilon}(u) = \frac{1}{2}\mathbb{E}[D(\xi, T)] + \frac{\alpha}{2} \int_{17}^{90} \|u(t)\|_2^2 dt + \frac{\gamma}{2}\mathbb{E}\left[\left|g_\varepsilon(\mathcal{R}(\xi, 90) - 1)\right|^2\right]. \tag{6.20}$$

Since the control is applied nonlinearly in the model, computation of derivatives of the cost function (6.20) is complicated. Thus, instead of the Newton method, we use the projected gradient descent method, where the gradient of (6.20) is calculated using finite differencing with anisotropic step sizes $10^{-6} \cdot \max(|u|, 0.1)$. The ODE (6.14) is solved using an implicit Euler method with a time step 0.1. In this experiment, we use a fixed Moreau-Yosida parameter $\gamma = 5 \cdot 10^5$ in all iterations, and the smoothing width is chosen as $\varepsilon = 50/\sqrt{\gamma}$. The iteration is stopped when the cost value does not decrease in two consecutive iterations. Each random variable (6.17) is discretized with $n = 3$ Gauss-Legendre quadrature nodes, and the TT approximations are carried out with a relative error tolerance of 10^{-2}. The control $u(t)$ is discretized using 7 Gauss-Legendre nodes on $[17, 90]$ with a Lagrangian interpolation in between.

In Figure 6, we compare optimizations without constraining $\mathcal{R}(\xi, 90)$ (left), and with the a.s. constraint (right) as described above. We plot the time evolution of the mean and confidence interval of the total number of hospitalized individuals, $I^C(t) = I^{C1}(t) + I^{C2}(t)$.

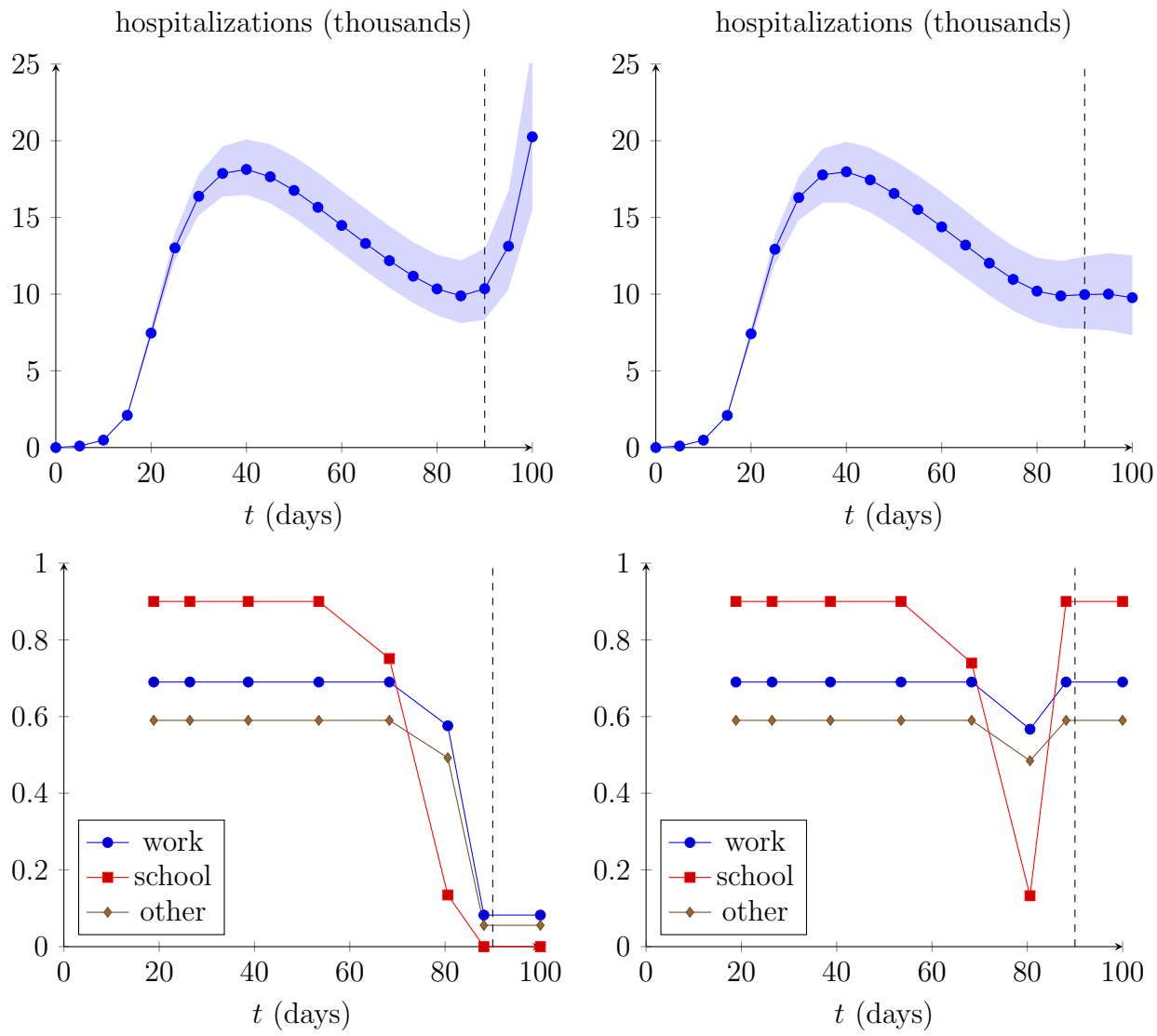

FIGURE 6. Top: optimized $I^C = I^{C1} + I^{C2}$, mean (blue circles) and 95% confidence interval (shaded area). Bottom: optimized control signals. Left: unconstrained optimization, Right: optimization constrained with $\mathcal{R}(\xi, 90) \leq 1$ a.s. approximated with $\gamma = 5 \cdot 10^5$. Black dashed lines indicate the end of the optimization time horizon $t = 90$.

The unconstrained scenario is a finite horizon optimization problem, which drives the control to near zero values at the end of the controllable time interval, $t = 90$, due to the zero terminal condition on the adjoint state. Naturally, this leads to infection growing again for $t > 90$, since we extrapolate these small values of the control from $t = 90$ onwards.

In contrast, if we constrain the $\mathcal{R}$ number at the end of the optimization interval to be below 1 almost surely, this drives the control to higher values again. If we extrapolate these control values beyond the optimization window, the epidemic continues decaying, albeit with

a slightly larger uncertainty. This indicates that almost sure constraints can suggest a more resilient control in risk-critical applications.

Appendix A. Proof of Lemma 3.11

Introduce a new variable $t = \exp(s/\varepsilon)$, then

$$\begin{aligned}\int_{-\infty}^{0} s\log(1+\exp(s/\varepsilon))ds &= \int_0^1 \frac{\varepsilon\log(t)\log(1+t)}{t/\varepsilon}dt \\ &= \varepsilon^2\int_0^1 \log(t)\log(t+1)d\log(t) \\ &= \frac{\varepsilon^2}{2}\left(\log(t)\right)^2\log(t+1)\Big|_0^1 - \frac{\varepsilon^2}{2}\int_0^1 (\log(t))^2 d\log(t+1).\end{aligned}$$

The first term is zero at $t=1$, and at $t=0$ we can use that $0 \le \log(t+1) \le t$ for $0 \le t < 1$ and $\lim_{t\to 0}(\log(t))^2\log(t+1) \le \lim_{t\to 0}(\log(t))^2 t = 0$. For the second term, we proceed as follows,

$$\begin{aligned}\int_{-\infty}^{0} s\log(1+\exp(s/\varepsilon))ds &= -\frac{\varepsilon^2}{2}\int_0^1 \frac{(\log(t))^2}{t+1}dt \\ &\ge -\frac{\varepsilon^2}{2}\int_0^1 (\log(t))^2 dt \\ &= -\frac{\varepsilon^2}{2}\underbrace{t(\log(t))^2\big|_0^1}_{0} + \varepsilon^2\int_0^1 \log(t)dt \\ &= \varepsilon^2\, t\log t\big|_0^1 - \varepsilon^2\int_0^1 dt = -\varepsilon^2.\end{aligned}$$

The proof is completed by recalling that $sg_\varepsilon(s) = \varepsilon\cdot s\log(1+\exp(s/\varepsilon))$.

References

[1] A. Alphonse, C. Geiersbach, M. Hintermüller, and T. M. Surowiec. Risk-averse optimal control of random elliptic variational inequalities. arXiv preprint 2210.03425, 2022.

[2] H. Antil, T.S. Brown, D. Verma, and M. Warma. Optimal control of fractional PDEs with state and control constraints. *Pure Appl. Funct. Anal.*, 7(5):1533–1560, 2022.

[3] H. Antil, S. Dolgov, and A. Onwunta. Ttrisk: Tensor train decomposition algorithm for risk averse optimization. *Numerical Linear Algebra with Applications*, n/a(n/a):e2481.

[4] H. Antil, D.P. Kouri, M.-D. Lacasse, and D. Ridzal, editors. *Frontiers in PDE-constrained optimization*, volume 163 of *The IMA Volumes in Mathematics and its Applications*. Springer, New York, 2018. Papers based on the workshop held at the Institute for Mathematics and its Applications, Minneapolis, MN, June 6–10, 2016.

[5] D. Bigoni, A. P. Engsig-Karup, and Y. M. Marzouk. Spectral tensor-train decomposition. *SIAM J. Sci. Comput.*, 38(4):A2405–A2439, 2016.

[6] E. Casas. Control of an elliptic problem with pointwise state constraints. *SIAM J. Control Optim.*, 24(6):1309–1318, 1986.

[7] S. Dolgov, B. N. Khoromskij, A. Litvinenko, and H. G. Matthies. Polynomial Chaos Expansion of random coefficients and the solution of stochastic partial differential equations in the Tensor Train format. *SIAM J. Uncertainty Quantification*, 3(1):1109–1135, 2015.

[8] S. Dolgov and D. Savostyanov. Parallel cross interpolation for high–precision calculation of high–dimensional integrals. *Comput. Phys. Commun.*, 246:106869, 2020.
[9] S. V. Dolgov, B. N. Khoromskij, I. V. Oseledets, and D. V. Savostyanov. Computation of extreme eigenvalues in higher dimensions using block tensor train format. *Comput. Phys. Commun.*, 185(4):1207–1216, 2014.
[10] S. V. Dolgov and D. V. Savostyanov. Alternating minimal energy methods for linear systems in higher dimensions. *SIAM Journal on Scientific Computing*, 36(5):A2248–A2271, 2014.
[11] R. Dutta, S. N. Gomes, D. Kalise, and L. Pacchiardi. Using mobility data in the design of optimal lockdown strategies for the COVID-19 pandemic. *PLoS Comput. Biol.*, 17(8):1–25, 2021.
[12] M. H. Farshbaf-Shaker, R. Henrion, and D. Hömberg. Properties of chance constraints in infinite dimensions with an application to PDE constrained optimization. *Set-Valued Var. Anal.*, 26(4):821–841, 2018.
[13] D.B. Gahururu, M. Hintermüller, and T.M. Surowiec. Risk-neutral pde-constrained generalized nash equilibrium problems. *Mathematical Programming*, 2022.
[14] 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(1):1–29, 2021.
[15] C. Geiersbach and W. Wollner. Optimality conditions for convex stochastic optimization problems in Banach spaces with almost sure state constraints. *SIAM J. Optim.*, 31(4):2455–2480, 2021.
[16] Caroline Geiersbach and Michael Hintermüller. Optimality Conditions and Moreau–Yosida Regularization for Almost Sure State Constraints. *ESAIM Control Optim. Calc. Var.*, 28:Paper No. 80, 36, 2022.
[17] A. Geletu, A. Hoffmann, P. Schmidt, and P. Li. Chance constrained optimization of elliptic PDE systems with a smoothing convex approximation. *ESAIM Control Optim. Calc. Var.*, 26:Paper No. 70, 28, 2020.
[18] S. A. Goreinov, I. V. Oseledets, D. V. Savostyanov, E. E. Tyrtyshnikov, and N. L. Zamarashkin. How to find a good submatrix. In V. Olshevsky and E. Tyrtyshnikov, editors, *Matrix Methods: Theory, Algorithms, Applications*, pages 247–256. World Scientific, Hackensack, NY, 2010.
[19] A. Gorodetsky, S. Karaman, and Y. Marzouk. A continuous analogue of the tensor-train decomposition. *Comput. Methods Appl. Mech. Engrg.*, 347:59–84, 2019.
[20] W. Hackbusch and B. N. Khoromskij. Low-rank Kronecker-product approximation to multi-dimensional nonlocal operators. I. Separable approximation of multi-variate functions. *Computing*, 76(3-4):177–202, 2006.
[21] M. Hintermüller and M. Hinze. Moreau-Yosida regularization in state constrained elliptic control problems: Error estimates and parameter adjustment. *SIAM Journal on Numerical Analysis*, 47(3):1666–1683, 2009.
[22] M. Hoffhues, W. Römisch, and T. M. Surowiec. On quantitative stability in infinite-dimensional optimization under uncertainty. *Optimization Letters*, 15(8):2733–2756, 2021.
[23] D. P. Kouri and T. M. Surowiec. Risk-averse PDE-constrained optimization using the conditional value-at-risk. *SIAM J. Optim.*, 26(1):365–396, 2016.
[24] K. Kunisch and D. Wachsmuth. Sufficient optimality conditions and semi-smooth Newton methods for optimal control of stationary variational inequalities. *ESAIM Control Optim. Calc. Var.*, 18(2):520–547, 2012.
[25] R. Löhner, H. Antil, S. Idelsohn, and E. Oñate. Detailed simulation of viral propagation in the built environment. *Comput. Mech.*, 66(5):1093–1107, 2020.
[26] R. Löhner, H. Antil, A. Srinivasan, S Idelsohn, and E. Oñate. High-fidelity simulation of pathogen propagation, transmission and mitigation in the built environment. *Archives of Computational Methods in Engineering*, pages 1–26, 2021.
[27] G. J. Lord, C. E. Powell, and T. Shardlow. *An Introduction to Computational Stochastic PDEs*. West Nyack: Cambridge University Press, 2014.
[28] K. Maute. Topology optimization under uncertainty. In *Topology optimization in structural and continuum mechanics*, pages 457–471. Springer, 2014.

[29] A. Yu. Mikhalev and I. V. Oseledets. Rectangular maximum–volume submatrices and their applications. *Linear Algebra Appl.*, 538:187–211, 2018.
[30] I. V. Oseledets. Tensor train decomposition. *SIAM J. Sci. Comp.*, 33(5):2295 – 2317, 2011.
[31] I. V. Oseledets and E. E. Tyrtyshnikov. TT-cross approximation for multidimensional arrays. *Linear Algebra Appl.*, 432(1):70–88, 2010.
[32] R. T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *The Journal of Risk*, 2:21 – 41, 2000.
[33] P. B. Rohrbach, S. Dolgov, L. Grasedyck, and R. Scheichl. Rank bounds for approximating Gaussian densities in the Tensor-Train format. *SIAM/ASA Journal on Uncertainty Quantification*, 10(3):1191–1224, 2022.
[34] D. V. Savostyanov and I. V. Oseledets. Fast adaptive interpolation of multi-dimensional arrays in tensor train format. In *Proceedings of 7th International Workshop on Multidimensional Systems (nDS)*. IEEE, 2011.
[35] R. Schneider and A. Uschmajew. Approximation rates for the hierarchical tensor format in periodic Sobolev spaces. *J. Complexity*, 2013.
[36] J. Sokołowski and J. P. Zolésio. *Introduction to shape optimization*, volume 16 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 1992. Shape sensitivity analysis.
[37] L. N. Trefethen. *Spectral methods in MATLAB*. SIAM, Philadelphia, 2000.
[38] F. Tröltzsch. *Optimal Control of Partial Differential Equations: Theory, Methods and Applications*. American Mathematical Society, 2010.

Harbir Antil, The Center for Mathematics and Artificial Intelligence (CMAI) and Department of Mathematical Sciences, George Mason University, Fairfax, VA 22030, USA.
Email address: hantil@gmu.edu

Sergey Dolgov, Department of Mathematical Sciences, University of Bath, Bath, BA2 7AY, UK.
Email address: s.dolgov@bath.ac.uk

Akwum Onwunta, Department of Industrial and Systems Engineering, Lehigh University, Bethlehem, PA 18015, USA.
Email address: ako221@lehigh.edu