Technische Universität München
Fakultät für Mathematik
Lehrstuhl für Mathematische Optimierung

Topics in PDE-Constrained Optimization under Uncertainty and Uncertainty Quantification

Johannes Milz

Vollständiger Abdruck der von der Fakultät für Mathematik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender:		Prof. Dr. Rainer Callies
Prüfer der Dissertation:	1.	Prof. Dr. Michael Ulbrich
	2.	Prof. Dr. Karl Kunisch
	3.	Prof. Dr. Alexander Shapiro

Die Dissertation wurde am 18.01.2021 bei der Technischen Universität München eingereicht und durch die Fakultät für Mathematik am 14.05.2021 angenommen.

Abstract

We develop an efficient sampling-free approximation scheme for moment-based distributionally robust nonlinear optimization problems. Our approach utilizes a smoothing method that allows the use of gradient-based optimization methods. We apply our scheme to finite-dimensional optimization problems and to optimal control problems with nonlinear partial differential equations. Furthermore, we apply the sample average approximation method to convex risk-neutral optimal control problems posed in Hilbert spaces and derive non-asymptotic error bounds, including exponential tail bounds, for their optimal controls and optimal values. Finally, we establish large deviations for the multilevel Monte Carlo mean estimator.

Zusammenfassung

Der Großteil der Arbeit befasst sich mit der Analyse und numerischen Umsetzung eines effizienten Ansatzes zur Approximation von momentenbasierten verteilungsrobusten nichtlinearen Optimierungsproblemen. Wir behandeln sowohl endlich-dimensionale Probleme als auch Steuerungsprobleme, die sich durch nichtlineare partiellen Differentialgleichungen ergeben. Desweiteren approximieren wir konvexe risikoneutrale Optimalsteuerungsprobleme, die in Hilberträumen gestellt sind, mittels empirischer Mittelwerte und leiten nicht-asymptotische Fehlerabschätzungen für optimale Steuerungen und Optimalwerte her. Im letzten Kapitel leiten wir große Abweichungen für Multilevel-Monte-Carlo-Schätzer her.

Contents

List of Prior Publications and Manuscripts

[234] J. Milz and M. Ulbrich, *An approximation scheme for distributionally robust nonlinear optimization*, SIAM J. Optim., 30 (2020), pp. 1996–2025, https://doi.org/10.1137/19M1263121.

[235] J. Milz and M. Ulbrich, *An approximation scheme for distributionally robust PDE-constrained optimization*, Preprint No. IGDK-2020-09. Technische Universität München, München, Jun. 2020, in review, http://www.igdk.eu/foswiki/pub/IGDK1754/Preprints/MilzUlbrich-PDEDRO.pdf.

Basic Notation and Preliminaries

General Sets

$\mathbb{N}_0$	$\mathbb{N} \cup \{0\}$
$\mathbb{R}_+$, $\mathbb{R}_{++}$	$[0,\infty)$, $(0,\infty)$
$\bar{\mathbb{R}}$	$\mathbb{R} \cup \{\pm\infty\}$
$\mathcal{D}$	bounded domain, $\mathcal{D} \subset \mathbb{R}^d$
$\operatorname{conv} A$	convex hull of A
$B_\epsilon(x)$	$\{\, y \in V : \|x-y\|_V < \epsilon \,\}$ for $x \in V$
$\bar{A}$ ($\bar{A}^{\|\cdot\|_V}$)	($\|\cdot\|_V$-)closure of S
$\|A\|$ ($\|\alpha\|$)	cardinality of A (length of multiindex $\alpha \in \mathbb{N}_0^d$)
$\dim(V)$	dimension of V
$\operatorname{span}\{\, f_k : k = 1,\dots,K \,\}$	span of $f_1,\dots,f_K$

General Notation

(x^k), (x_k)	sequences
$(x^k)_K$, $(x_k)_K$	subsequences
$(x^k,y^k)_{\mathbb{N}_0}$, $(x_k,y_k)_{\mathbb{N}_0}$	$((x^k,y^k))$, $((x_k,y_k))$
$(x^k,y^k)_K$, $(x_k,y_k)_K$	$((x^k,y^k))_K$, $((x_k,y_k))_K$
$(\cdot)_+$	$\max\{0,\cdot\}$ (componentwise)
$\|\cdot\|_p$	vector p-norm ($1 \le p \le \infty$)
$\langle\cdot,\cdot\rangle_{V^*,V}$	duality paring
1_A	indicator function of A
$f : V_1 \rightrightarrows V_2$	set-valued mapping, multifunction
A^*	(Hilbert space-)adjoint operator of A

Derivatives, Gradients and Subdifferentials

∇f ($\nabla_x f$)	(partial) gradient of f
$\nabla_{xx} f$	partial Hessian of f
$\mathrm{D} f$	Gâteaux, Hadamard or Fréchet derivative of f
$\mathrm{D}_x f$, f_x	partial Gâteaux or Fréchet derivative of f
$f'(x;h)$	directional derivative of f at x in the direction h
∂f	(convex) subdifferential of f, Clarke's generalized gradient of f
$\mathrm{D}^\alpha f$	weak derivative of f of order α, $\alpha \in \mathbb{N}_0^d$ is a multiindex

Matrices

I	identity matrix
$\cdot \bullet \cdot$	Frobenius inner-product
$\mathrm{Diag}(a)$	diagonal matrix with $\mathrm{Diag}(a)_{ii} = a_i$
$\mathbb{S}^p$, $(\mathbb{S}^p_+)$, $[\mathbb{S}^p_{++}]$	set of sym. (positive semidefinite) [positive definite] $p \times p$ matrices
$\preccurlyeq$	Löwner partial order
$\lambda_{\max}(A)$ $(\lambda_{\min}(A))$	maximum (minimum) eigenvalue of $A \in \mathbb{S}^p$
$\lambda(A)$	eigenvalues of $A \in \mathbb{S}^p$ with $\lambda_1(A) \geq \cdots \geq \lambda_p(A)$
$\lambda : \mathbb{S}^p \to \mathbb{R}^p$	eigenvalue mapping
$A^{1/2}$	square root of $A \in \mathbb{S}^p_{++}$
$\|\cdot\|_2$	spectral norm
$\|\cdot\|_A$	$\|A^{1/2} \cdot\|_2$ for $A \in \mathbb{S}^p_{++}$
$N(A)$	null space of A
A^+	Moore–Penrose inverse of A

Probability and Measure

$(\Omega, \mathcal{F}, P)$	probability space
$\mathcal{M}$	set of probability measures on $\mathbb{R}^p$
$\mathbb{E}$ $(\mathbb{E}_P)$	expectation (w.r.t. $P \in \mathcal{M}$)
$\mathrm{E}^N[Z]$	sample mean, $(1/N)\sum_{i=1}^N Z_i$
Cov (Cov_P)	covariance (w.r.t. $P \in \mathcal{M}$)
$\mathcal{N}(\mu, \Sigma)$	normal distribution with mean μ and covariance Σ
$\mathcal{B}(V)$	Borel-σ-field of V
$\mathrm{Prob}(A)$	probability of event A
$(T, \mathcal{A}, \nu)$	measurable space
$L^0(T; V)$	$L^0(T, \mathcal{A}, \nu; V)$, class of strongly ($\nu$-)measurable functions $f : T \to V$

Normed and Banach Spaces

$\mathscr{L}(V_1, V_2)$	space of bounded, linear operators from V_1 to space V_2, equipped with $\|\cdot\|_{\mathscr{L}(V_1,V_2)} = \sup_{v \in V_1, \|v\|_{V_1}=1} \|\cdot v\|_{V_2}$
V^*	$\mathscr{L}(V, \mathbb{R})$
$V_1 \times V_2$	Cartesian product, equipped with $\|\cdot\|_{V_1 \times V_2} = (\|\cdot\|^2_{V_1} + \|\cdot\|^2_{V_2})^{1/2}$
$L^p(T; V)$	space of $Z \in L^0(T; V)$ with $\|Z\|_{L^p(T;V)} = (\int_T Z(t) \mathrm{d}\nu(t))^{1/p} < \infty$ $(1 \leq p < \infty)$
$L^\infty(T; V)$	space of $Z \in L^0(T; V)$ with $\|Z\|_{L^\infty(T;V)} = \operatorname{ess\,sup}_{t \in T} \|Z(t)\|_V < \infty$
$L^p(T)$	$L^p(T; \mathbb{R})$ $(1 \leq p \leq \infty)$
$W^{s,p}(\mathcal{D})$	$\{ u \in L^p(\mathcal{D}) : \mathrm{D}^\alpha u \in L^p(\mathcal{D}),\ \|\alpha\| \leq s \}$, Sobolev space $(s \in \mathbb{N}_0,\ 1 \leq p < \infty)$, equipped with $\|\cdot\|_{W^{s,p}(\mathcal{D})} = (\sum_{\|\alpha\| \leq s} \|\mathrm{D}^\alpha \cdot\|^p_{L^p(\mathcal{D})})^{1/p}$
$H^s(\mathcal{D})$	$W^{s,2}(\mathcal{D})$
$\|\cdot\|_{H^s(\mathcal{D})}$	seminorm on $H^s(\mathcal{D})$, $(\sum_{\|\alpha\|=s} \|\mathrm{D}^\alpha \cdot\|^2_{L^2(\mathcal{D})})^{1/2}$ $(s \in \mathbb{N})$
$H^1_0(\mathcal{D})$	$\|\cdot\|_{W^{1,2}(\mathcal{D})}$-closure of $C^\infty_0(\mathcal{D})$, equipped with $\|\cdot\|_{H^1_0(\mathcal{D})} = \|\cdot\|_{H^1(\mathcal{D})}$

For each normed space, the underlying field is $\mathbb{R}$. For a locally Lipschitz continuous function $h : \mathbb{R}^n \to \mathbb{R}$, the Clarke subdifferential $\partial h(x)$ at $x \in \mathbb{R}^n$ is considered a subset of $\mathbb{R}^n$. Let $\iota : V_1 \to V_2$ be the embedding operator of the continuous embedding $V_1 \hookrightarrow V_2$ [46, Def. 6.1]. We identify V_1 with $\iota(V_1)$, and write $v \in V_2$ instead of $\iota v \in V_2$ for $v \in V_1$. The dual of $V_1 \times V_2$ is often identified with $V_1^* \times V_2^*$. Here, V_1 and V_2 are normed spaces. For a normed space $(V, \|\cdot\|_V)$, $\|\cdot\|_{V^*}$ is called the *dual norm* to $\|\cdot\|_V$. We use the fact that $\|f\|_{V^*} = \sup_{\|v\|_V \leq 1} \langle f, v \rangle_{V^*,V}$ for $f \in V^*$ [211, p. 75].

A normed space is endowed with its Borel-σ-field if not specified otherwise. Unless stated otherwise, all relations between random variables hold w.p. 1 (with probability one). We use ξ to denote a measurable mapping $\xi : \Omega \to \Xi$ as well as a deterministic element $\xi \in \Xi$. A function $Z \in L^0(T; V)$ is called *Bochner/strongly measurable*, and $Z \in L^1(T; V)$ is referred to as *Bochner/strongly integrable*. A random variable $\xi : \Omega \to \mathbb{R}$ is *sub-Gaussian with parameter* τ if $\tau \geq 0$ and $\mathbb{E}[\exp(\lambda\xi)] \leq \exp(\lambda^2\tau^2/2)$ for all $\lambda \in \mathbb{R}$ [57, p. 2]. For example, a centered Gaussian random variable with variance σ^2 is sub-Gaussian with parameter σ [57, p. 9].
The notion of a probability space is defined, for instance, in [39, p. 25]. The definition of the Bochner spaces and the Borel-σ-algebra is provided, for example, in [159, pp. 2 and 21]. The Lebesgue spaces $L^p(\mathcal{D})$, the Sobolev spaces $W^{s,p}(\mathcal{D})$ and $H_0^1(\mathcal{D})$, and weak derivatives are defined in [1, pp. 21–22 and 44–45], for example. Since $\mathcal{D} \subset \mathbb{R}^d$ is a bounded domain, $(H_0^1(\mathcal{D}), \|\cdot\|_{H_0^1(\mathcal{D})})$ is a Hilbert space [151, pp. 21–22]. Clarke's generalized gradient is defined in [79, p. 10], and the (convex) subdifferential in [46, p. 81].

Introduction

Parameters in physics-based models may be uncertain, such as diffusion coefficients in partial differential equations (PDEs). Such uncertainty can arise from a lack of knowledge about the process being modeled or an inherent variability of the model's parameters [132].
In the field of (PDE-constrained) optimization under uncertainty, several approaches have been proposed for obtaining decisions that are resilient to uncertainty. When uncertain parameters are modeled as a random vector, such approaches include risk-neutral, risk-averse and distributionally robust (stochastic) optimization [190, 186, 269, 292].
If the random vector's distribution is known, we can formulate a risk-neutral optimization problem, that is, the minimization of the expected value of a parameterized objective function where the expectation is taken w.r.t. the known probability distribution [292].
The framework of distributionally robust optimization allows for incomplete knowledge about the parameter vector's distribution. For example, while the distribution of the parameter vector may be unknown, its first and second centered moment may be available [94, 283], or its distribution is known to be contained in some ball about a known reference probability measure [107, 119]. The knowledge about the parameter vector's distribution is collected in a set of probability measures, the ambiguity set. A distributionally robust optimization problem is formulated as the minimization of the worst-case expected value of the parameterized objective function, where the worst-case is computed w.r.t. all probability measures contained in the ambiguity set.
The dissertation's main focus is on the development of an effective sampling-free approximation scheme for distributionally robust nonlinear optimization problems where the ambiguity set is defined by conditions on the parameter vector's moments. Furthermore, we provide non-asymptotic performance guarantees for the sample average approximation method—an approach for approximating risk-neutral optimization problems—applied to stochastic convex optimal control problems. Finally, we develop a further analysis of the Multilevel Monte Carlo mean estimator that complements the mean-squared error analysis available in the literature [37, 134]. In the following, we provide a more detailed introduction and overview of the topics covered in the dissertation and outline the main contributions made.

An Introduction to Distributionally Robust Optimization. Stochastic programming offers a methodology for optimization under uncertainty. Its application requires the uncertain parameters to be modeled as a random vector distributed according to a probability distribution. Distributionally robust optimization (DRO) (also called distributionally robust stochastic optimization) is a framework that allows for incomplete knowledge about the parameter vector's distribution with the goal of computing an optimal solution that is resilient to the distributional uncertainty. For example, the unknown parameter vectors' distribution may be approximated by a known nominal probability measure defined by, for instance, historical data. In this case, DRO allows the decision maker to incorporate the "uncertainty" of the nominal probability distribution [94, 300]. The task is formulated as a min-max problem—the minimization of the worst-case expected value of a parameterized objective function.
The maximization problem's feasible set is defined by the probability measures contained in an ambiguity set. This set models the distributional uncertainty and can be defined through moment constraints [283, 94, 344, 264, 136] and/or various probability distances [119, 253, 293, 107, 120]. Popular choices for such probability distances are: the Wasserstein distance [119, 299,

43, 359, 107], the Prokhorov metric [106], and the ϕ-divergence [100, 293, 340, 21]. When the ambiguity set is defined by moment constraints, we refer to the resulting DRO problem (DROP) as a moment-based DROP.

A typical solution approach for certain moment-based DROPs exploits Lagrangian duality [94, 344, 75, 346]. For example, if the ambiguity set is conic representable, then the maximization problems are conic linear programs [344]. Under mild conditions, strong duality is satisfied and the Lagrangian dual of the linear program can be concatenated with the upper-level problem to obtain an equivalent reformulation of the DROP as a single-level problem with linear matrix inequalities as constraints [94, 344, 290].

The tractability of the maximization problem's dual depends on the structure of the parameterized objective function [94, 344, 23]. For example, if this function, as a mapping of the parameters, is the pointwise maximum of affine functions or quadratic functions and the ambiguity set is conic representable, then the dual is tractable (see [94, sect. 4.1], [344, sect. 2], and Proposition 2 in the supplementary material accompanying [344]). Without such structural properties, the maximization problems with the ambiguity set as the feasible set may be intractable [94, 344].

The moment-based DROPs considered in this dissertation differ from those in [94, 344, 75] in that, for example, the parameterized objective functions, as mappings of the parameters, are generally non-quadratic and nonlinear. Moreover, the DROPs involving PDEs are defined by parameterized objective functions that are, in addition to being non-quadratic and nonlinear, implicitly defined by the solution operators of the parameterized PDEs. For these DROPs, the solution approach developed in [94, 344, 75] does not result in single-level programs with explicit representations of their objective and constraint functions that allow the application of available PDE-constrained optimization methods.

DRO is one approach to optimization under uncertainty. Further methodologies for optimization under uncertainty are: risk-neutral optimization [294], risk-averse optimization [269, 294], (ambiguous) chance-constrained programming [245, 106, 354, 361, 63], and robust optimization [24, 23, 30, 137, 206]. DRO has several links to these approaches. For example, if the ambiguity set is a singleton, then a DROP is a risk-neutral problem, and if the ambiguity set consists of all probability measures supported on some (compact) set, then a DROP becomes a robust optimization problem [295, 23]. Moreover, risk-averse optimization problems with coherent risk measures can be equivalently reformulated as min-max problems where the maximum is taken over the domain of the convex conjugate of the risk measure [269, 274, 294].

Approximation Scheme for Distributionally Robust Nonlinear Optimization. In Chapter 1, we consider the moment-based distributionally robust nonlinear optimization problem

$$\min_{x\in\mathbb{R}^n} \sup_{P\in\mathcal{P}} \mathbb{E}_P[f_0(x,\xi)] \quad \text{s.t.} \quad \sup_{P\in\mathcal{P}} \mathbb{E}_P[f_k(x,\xi)] \leq 0, \quad k \in K \setminus \{0\}, \tag{1}$$

and develop an approximation scheme and a solution approach for it. Here, $f_k : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}$ are the parametrized functions, and $\{0\} \subset K \subset \mathbb{N}_0$ with $|K| < \infty$. The moment-based ambiguity set $\mathcal{P}$ is defined in (1.1.2) and built on those considered in [94, 344, 75]. It ensures that the DROP (1), defined by the parametrized functions f_k with nonlinear and non-quadratic dependence on the parameters, is well-posed under mild conditions on f_k while retaining a statistical interpretation and enabling a data-driven definition (see section 1.8.2). In contrast to the above motivation of DRO, we also allow distributionally robust constraints in (1).

Instead of using Lagrangian duality to reformulate the DROP (1) as in [94, 344, 75], we use second-order expansions of the parameterized functions w.r.t. the parameters, allowing us to compute the expected value of the surrogate functions explicitly. For $x \in \mathbb{R}^n$, we approximate

$f_k(x,\cdot)$ using a quadratic surrogate $m_k(x,\cdot)$, formally defined in (1.1.3), and we formulate the approximated DROP as

$$\min_{x\in\mathbb{R}^n} \sup_{P\in\mathcal{P}} \mathbb{E}_P[m_0(x,\xi)] \quad \text{s.t.} \quad \sup_{P\in\mathcal{P}} \mathbb{E}_P[m_k(x,\xi)] \leq 0, \quad k \in K\setminus\{0\}.$$

Since $m_k(x,\cdot)$ is quadratic, the definition of the ambiguity set $\mathcal{P}$ will allow us to explicitly compute $\sup_{P\in\mathcal{P}} \mathbb{E}_P[m_k(x,\xi)]$ for $x\in\mathbb{R}^n$. For each $k\in K$, the function $\sup_{P\in\mathcal{P}} \mathbb{E}_P[m_k(\cdot,\xi)]$ is the sum of the optimal value functions defined by a nonconvex trust-region problem and a semidefinite program. These optimal value functions can be efficiently evaluated, as opposed to those in (1). However, they are generally nonsmooth. For these optimal value functions, we construct smoothing functions, that is, smooth approximations with explicit bounds on the smoothing error, exploiting the strong Lagrangian duality for the trust-region problems and a solution formula for the semidefinite programs. Moreover, we demonstrate that the smoothing functions can efficiently be evaluated and that they satisfy gradient consistency.
For the numerical solution of the approximated DROPs, we develop a smoothing method that computes a sequence of (approximate) stationary points of smoothed DROPs, defined by the smoothing functions, while it decreases smoothing parameters to zero. We prove the convergence of the sequence generated by the smoothing method towards stationary points of the approximated DROP exploiting the gradient consistency of the smoothing functions. Moreover, we show that the approximated DROP can be equivalently formulated as a nonlinear semidefinite program. We compare our algorithmic approach with the application of the solver `PENLAB` [110] for nonlinear semidefinite programs and the proximal bundle method `MPBNGC` [223, 224]. For the proximal bundle method, we implemented the `Julia` interface `MPBNGCInterface.jl`.
Chapter 1 is based on the article [234].

Approximation Scheme for Distributionally Robust PDE-Constrained Optimization. In Chapter 2, we extend our approximation scheme and solution approach to the distributionally robust PDE-constrained optimal control problem

$$\min_{u\in U_{\text{ad}}} \Big\{ \sup_{P\in\mathcal{P}} \mathbb{E}_P[J(S(u,\xi),u,\xi)] \Big\}, \tag{2}$$

where the set of admissible controls U_{ad} is a subset of the Hilbert space U. Moreover, $J : Y\times U\times\mathbb{R}^p \to \mathbb{R}$ is the parametrized objective function, $S : U_{\text{ad}}\times\mathbb{R}^p \to Y$ is the solution operator of a parameterized nonlinear PDE, and Y is a Banach space. The ambiguity set $\mathcal{P}$ is defined in (2.1.2), motivated by the results established in section 1.8.2.
The surrogate function, which we use to approximate the DROP (2), is defined by a second-order Taylor's expansion of the function $\xi \mapsto J(S(u,\xi),u,\xi)$. The smoothing functions constructed in Chapter 1 are used to define smoothed DROPs. We prove the existence of optimal solutions for the DROP (2), and the associated approximated and smoothed DROPs. The ambiguity set defined in (2.1.2) is weakly-star sequentially compact, allowing us to establish the existence of a worst-case distribution of the maximization problem in (2).
We extend the smoothing method developed in Chapter 1 to allow the numerical treatment of the approximated DROPs posed in Hilbert spaces using existing, derivative-based solvers for PDE-constrained optimization. We show that a sequence of global solutions, generated by the smoothing method, converges towards an optimal solution of the approximated DROP. In order to evaluate the surrogate function as well as the smoothing functions and their derivatives, we use the adjoint approach.
We present numerical results for the DRO of the steady Burgers' equation and of the unsteady Burgers' equation. The Burgers' equation is a one-dimensional PDE that models convection-diffusion phenomena, such as shock waves and supersonic flow [92, p. 203], [317, p. 649].

Moreover, we provide conditions sufficient to ensure the weak-weak continuity of solution operators of PDEs.
Chapter 2 is based on the manuscript [235].

Sample Average Approximation for Stochastic Convex Optimal Control Problems: Non-Asymptotic Sample Size Estimates. In Chapter 3, we investigate performance guarantees for the sample average approximation (SAA) approach applied to the stochastic optimal control problem

$$\min_{u \in U_{\text{ad}}} \{ f(u) = \mathbb{E}[\widehat{J}(u, \xi)] + \Psi(u) \}, \tag{3}$$

where U_{ad} is a convex subset of the Hilbert space U, $\xi : \Omega \to \Xi$ is a random vector, and $\Psi : U_{\text{ad}} \to \mathbb{R} \cup \{\infty\}$ is convex. Moreover, $\widehat{J} : U \times \Xi \to \mathbb{R}$ is a Carathéodory function and $\widehat{J}(\cdot, \xi)$ is α-strongly convex for all $\xi \in \Xi$ and some $\alpha \geq 0$.
The SAA method approximates the expected value in (3) using the sample average computed with N independent samples of ξ, thereby defining the SAA problem [291]. Our main focus is on deriving the following exponential tail bound for the distance between an optimal solution u^* to (3) and a minimizer u_N^* to its SAA problem (under additional assumptions on $\widehat{J}$ and $\alpha > 0$):

$$\text{Prob}(\|u^* - u_N^*\|_U \geq \varepsilon) \leq 2 \exp(-\tau^{-2} N \varepsilon^2 \alpha^2 / 3) \quad \text{for all} \quad \varepsilon > 0,$$

where $\tau > 0$ depends on certain properties of the parameterized objective function $\widehat{J}$.
The exponential tail bound yields the following non-asymptotic sample size estimate: if $\varepsilon > 0$, $\delta \in (0, 1)$ and $N \geq 3 \ln(2/\delta)(\tau/\varepsilon\alpha)^2$, then $\|u^* - u_N^*\|_U < \varepsilon$ with a probability of at least $1 - \delta$.
Chapter 3 also offers the non-asymptotic analysis of the SAA problem's optimal value, allowing us to derive non-asymptotic confidence intervals for the optimal value of (3).
We demonstrate that our assumptions are fulfilled for a class of linear-quadratic optimal control problems governed by parameterized affine-linear PDEs with essentially bounded random inputs; a problem class that has extensively been investigated in the literature [125, 226, 230].
The SAA problem for (3) is an infinite-dimensional optimization problem. In section 3.3, we discretize an instance of this SAA problem using finite elements, and we derive reliable error bounds on the distance between the optimal solution of the discretized SAA problem and the minimizer of the risk-neutral control problem (3).
Finally, we analyze the expected value of the SAA problem's optimal value for risk-averse convex control problems using the superquantile/conditional value-at-risk.

Exponential Tail Bounds for Multilevel Monte Carlo Mean Estimators in a Class of Smooth Banach Spaces. In Chapter 4, we derive exponential tail bounds for the Multilevel Monte Carlo (MLMC) mean estimator. To estimate the mean of solutions to PDEs with random inputs, the MLMC mean estimator has been identified as an effective method, as it can exploit low and high fidelity PDE models resulting from, for example, finite element approximations [37, 134, 311]. MLMC methods utilize several low and high fidelity models and are designed to perform many simulations with the low fidelity models, but relatively few with accurate approximations.
We augment the available mean-squared error analysis [37, 134, 146] by deriving exponential tail bounds/large deviations for the MLMC mean estimator. These exponential tail bounds provide complementary performance guarantees for the MLMC mean estimator, but the analysis requires more restrictive assumptions on the models than that of the mean-squared errors.
The exponential tail bounds are derived for random variables that take values in certain smooth Banach spaces. Examples of such spaces are all Hilbert spaces and all Sobolev spaces of at least square integrable functions.

A Common Theme of the Dissertation. All chapters make use of the notions of sub-Gaussianity in one way or another. For example, the concept of sub-Gaussianity forms the basis for defining the ambiguity set $\mathcal{P}$ in section 1.1 and is used in section 1.6.4 to provide a data-driven definition of the ambiguity set. The existence of a worst-case distribution and the uniform integrability of the (reduced) parameterized objective function also rely on the properties of sub-Gaussian distributions as discussed in sections 2.3 and 2.7. We derive exponential tails bounds for the sums of independent, sub-Gaussian Hilbert and Banach space-valued random variables, and use them for analyzing the reliability and accuracy of the SAA problem's optimal solutions in Chapter 3 and of the MLMC mean estimator in Chapter 4.
Sub-Gaussianity and closely related concepts form the basis of non-asymptotic statistics [57, 337, 167] and are used in different fields of optimization, such as robust optimization [25], chance-constrained programming [244], risk-neutral and risk-averse optimization [243, 208], and distributionally robust optimization [94, 107, 300]. Further areas of application include compressed sensing [115, 167], scientific simulation [318], and numerical analysis [149].

Structure of the Dissertation. The dissertation is divided into four chapters, each of which focuses on a different topic. The topics in the first two chapters have a strong link, whereas the third and fourth chapters are mostly independent. Technical proofs and auxiliary results, if any, are presented at the end of each chapter. We provide a comprehensive literature review in each chapter. Basic notation and some preliminaries are summarized on pp. vii–ix.
For some calculations, such as the computation of integrals or solutions to boundary value problems, we adapt the approach used in [49, p. viii] and provide external links to Wolfram|Alpha that show these calculations. When clicking on the corresponding formula, the reader is redirected to Wolfram|Alpha. (Unfortunately, the links are unavailable in the printed version.) For example, $y(x) = x(1-x)$ for $x \in [0,1]$ is the solution to the boundary value problem $-y'' = 2$ in $(0,1)$ with $y(0) = y(1) = 0$.

1 Approximation Scheme for Distributionally Robust Nonlinear Optimization

We develop a sampling-free approximation scheme for distributionally robust optimization problems (DROPs) with nonlinear, nonconcave dependence on uncertain parameters. We define the ambiguity set through moment constraints. In order to make the computation of first-order stationary points computationally tractable, we approximate nonlinear functions using quadratic expansions with respect to the parameters, resulting in lower-level problems defined by trust-region problems and semidefinite programs. Subsequently, we construct smoothing functions for the approximated lower-level functions which can efficiently be evaluated. We use a smoothing method that computes a sequence of stationary points of the smoothed DROPs while smoothing parameters are decreased to zero, and establish the convergence to stationary points of the approximated DROP. For the numerical simulations, we construct twenty test problems from the Moré–Garbow–Hillstrom test set. We compare the performance of the smoothing method with a bundle method and a solver for nonlinear semidefinite optimization problems.
The chapter is mainly based on the article

[234] J. Milz and M. Ulbrich, *An approximation scheme for distributionally robust nonlinear optimization*, SIAM J. Optim., 30 (2020), pp. 1996–2025, `https://doi.org/10.1137/19M1263121`. First Published in SIAM Journal on Optimization in vol. 30 no. 3, published by the Society for Industrial and Applied Mathematics (SIAM), Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

The results of the work [234] are reproduced under the Author's Rights statute of the SIAM *Consent to Publish* agreement.
To be consistent with the notation used in Chapter 2, we use a slightly different definition of the optimal value functions defined by the trust-region problems (1.1.6) than in [234], and use a notion of a smoothing function that implies that used in [234, Def. 3.1]. To underpin the statements made in [234, pp. 2002 and 2019], we report numerical results comparing the performance of solvers for certain semidefinite programs with an implementation of the solution formulas provided by Proposition 1.3.1 in section 1.8.4. In section 1.6.1, we report the results on the numerical simulations announced in [234, p. 2019] where we used `Ipopt` [336] with a modified line search within the smoothing method. We provide a further discussion on the data-driven definition of the ambiguity set and some of its properties in section 1.8.2. In section 1.8.3, a proof of the equivalent formulation of the approximated DROP as a nonlinear semidefinite program used in [234] is provided.
Lemma 1.8.2 is taken from the manuscript

[235] J. Milz and M. Ulbrich, *An approximation scheme for distributionally robust PDE-constrained optimization*, Preprint No. IGDK-2020-09. Technische Universität München, München, Jun. 2020, in review, `http://www.igdk.eu/foswiki/pub/IGDK1754/Preprints/MilzUlbrich-PDEDRO.pdf`.

Parts of the numerical simulations were preformed using an earlier version of the `Julia` interface

J. Milz, *MPBNGCInterface.jl: A Julia package for interfacing the multiobjective proximal bundle method MPBNGC*, Technische Universität München, München, Mar. 2020, `https://github.com/milzj/MPBNGCInterface.jl`.

The current version provides a full interface to `MPBNGC` [223, 224], and has an extended set of examples and tests.

1.1 Introduction

Distributionally robust optimization (DRO) is a popular methodology used to obtain solutions to optimization problems that are resilient to distributional uncertainty [94, 107, 136, 293, 344, 264, 253]. We develop a sampling-free approximation scheme for the distributionally robust nonlinear optimization problem (DROP)

$$\min_{x\in\mathbb{R}^n} \sup_{P\in\mathcal{P}} \mathbb{E}_P[f_0(x,\xi)] \quad \text{s.t.} \quad \sup_{P\in\mathcal{P}} \mathbb{E}_P[f_j(x,\xi)] \le 0, \quad j \in J\setminus\{0\}, \tag{1.1.1}$$

where $f_j : \mathbb{R}^n\times\mathbb{R}^p \to \mathbb{R}$, and $\{0\} \subset J \subset \mathbb{N}_0$ with $|J| < \infty$. The ambiguity set $\mathcal{P}$ is defined through moment constraints and an entropic dominance constraint similar to those in [75, 94, 300]:

$$\begin{aligned}\mathcal{P} = \{P \in \mathcal{M} :\ & \|\bar{\Sigma}^{-1/2}(\mathbb{E}_P[\xi] - \bar{\mu})\|_2 \le \Delta, \quad \bar{\Sigma}_0 \preccurlyeq \operatorname{Cov}_P[\xi] \preccurlyeq \bar{\Sigma}_1, \\ & \ln \mathbb{E}_P[\exp\left(d^T(\xi - \mathbb{E}_P[\xi])\right)] \le (1/2)d^T\bar{\Sigma}_1 d \quad \text{for all} \quad d \in \mathbb{R}^p\,\},\end{aligned} \tag{1.1.2}$$

where $\Delta > 0$, $\bar{\mu} \in \mathbb{R}^p$, $\bar{\Sigma}_0$, $\bar{\Sigma}_1$, $\bar{\Sigma}_1 - \bar{\Sigma}_0 \in \mathbb{S}^p_+$, and $\bar{\Sigma} \in \mathbb{S}^p_{++}$. The first two conditions in (1.1.2) model confidence regions for the mean and the covariance of the random vector ξ. The ambiguity set $\mathcal{P}$ contains all distributions of strictly sub-Gaussian random vectors, in particular all normal distributions, with mean μ satisfying $\|\bar{\Sigma}^{-1/2}(\mu - \bar{\mu})\|_2 \le \Delta$ and covariance matrix Σ fulfilling $\sigma_0\bar{\Sigma} \preccurlyeq \Sigma \preccurlyeq \sigma_1\bar{\Sigma}$; see [57, pp. 185–186]. In section 1.8.2, we show that the data, such as $\bar{\mu}$ and $\bar{\Sigma}$, used in (1.1.2) may be defined using empirical estimates, similar to the choices made in [94, sect. 3.4] and [300, sect. 3.3]. The entropic dominance constraint in (1.1.2) implies that $\sup_{P\in\mathcal{P}} \mathbb{E}_P[f_j(\cdot,\xi)]$ is finite-valued under mild conditions on $f_j(x,\cdot)$ for $x \in \mathbb{R}^n$ (see section 2.3), and implicitly imposes higher-order moment constraints (see Lemma 1.8.2). We provide further details on the ambiguity set $\mathcal{P}$ in section 1.8.2.

To obtain tractable approximations of the objective function and each of the constraint functions in (1.1.1), we approximate $f_j(x,\cdot)$ using a second-order expansion $m_j(x,\cdot)$ defined by

$$m_j(x,\xi) = a_j(x) + b_j(x)^T(\xi - \bar{\mu}) + (1/2)(\xi - \bar{\mu})^T C_j(x)(\xi - \bar{\mu}), \tag{1.1.3}$$

where $m_j : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}$, $a_j : \mathbb{R}^n \to \mathbb{R}$, $b_j : \mathbb{R}^n \to \mathbb{R}^p$, and $C_j : \mathbb{R}^n \to \mathbb{S}^p$. If, for each $x \in \mathbb{R}^n$, $f_j(x,\cdot)$ is twice differentiable at $\bar{\mu}$, possible choices for a_j, b_j, and C_j are $a_j = f_j(\cdot,\bar{\mu})$, $b_j = \nabla_\xi f_j(\cdot,\bar{\mu})$, and $C_j = \nabla_{\xi\xi} f_j(\cdot,\bar{\mu})$, respectively. We formulate the approximated DROP

$$\min_{x\in\mathbb{R}^n} \sup_{P\in\mathcal{P}} \mathbb{E}_P[m_0(x,\xi)] \quad \text{s.t.} \quad \sup_{P\in\mathcal{P}} \mathbb{E}_P[m_j(x,\xi)] \le 0, \quad j \in J\setminus\{0\}. \tag{1.1.4}$$

We show that each lower-level optimization problem in (1.1.4) separates into the semidefinite program (SDP)

$$\varphi_j(x) = \max_{\Sigma\in\mathbb{S}^p} \left\{\, (1/2)C_j(x) \bullet \Sigma : \quad \bar{\Sigma}_0 \preccurlyeq \Sigma \preccurlyeq \bar{\Sigma}_1 \,\right\} \tag{1.1.5}$$

and the nonconvex trust-region problem (TRP)

$$\psi_j(x) = \max_{d\in\mathbb{R}^p} \left\{\, b_j(x)^T d + (1/2)d^T C_j(x) d : \quad \|\bar{\Sigma}^{-1/2}d\|_2 \le \Delta \,\right\}. \tag{1.1.6}$$

Here $\psi_j : \mathbb{R}^n \to \mathbb{R}$ and $\varphi_j : \mathbb{R}^n \to \mathbb{R}$. We have

$$\mathbb{E}_P[m_j(x,\xi)] = a_j(x) + b_j(x)^T d + (1/2) d^T C_j(x) d + (1/2) C_j(x) \bullet \mathrm{Cov}_P[\xi], \tag{1.1.7}$$

where $d = \mathbb{E}_P[\xi] - \bar{\mu}$; see, e.g., [42, Lem. 1.1.2]. Combined with the definition of the ambiguity set $\mathcal{P}$ provided in (1.1.2), and (1.1.7), we find that, for each $j \in J$,

$$\sup_{P \in \mathcal{P}} \mathbb{E}_P[m_j(x,\xi)] = a_j(x) + \varphi_j(x) + \psi_j(x).$$

Hence, the approximated DROP (1.1.4) is equivalent to

$$\min_{x \in \mathbb{R}^n} a_0(x) + \varphi_0(x) + \psi_0(x) \quad \text{s.t.} \quad a_j(x) + \varphi_j(x) + \psi_j(x) \leq 0, \quad j \in J \setminus \{0\}.$$

The optimal value functions (1.1.6) and (1.1.5), which are generally nonsmooth, provide tractable approximations of the lower-level problems in (1.1.1). We construct smoothing functions of them, and we use these functions to define smoothed DROPs. Using a smoothing method, which is similar to those developed in [72, 349], we compute a sequence of stationary points of the smoothed DROPs while decreasing smoothing parameters to zero. Our approach allows us to obtain Clarke stationary points of the approximated DROP (1.1.4).

In order to obtain a smoothing function for the optimal value function defined in (1.1.5), we use the fact that the SDP (1.1.5) can be solved analytically provided that the eigenvalues of a transformation of $C_j(x)$ are available [350, Thm. 2.2]. Combined with the theory on spectral functions established in [215, 319], we construct a smoothing function for (1.1.5). Besides the construction of a smoothing function, the solution formula allows for a significantly faster solution of the SDP (1.1.5) than state-of-the-art SDP solvers as we demonstrate in section 1.8.4. Our smoothing approach for the optimal value function of the TRP (1.1.6) utilizes strong duality for TRPs [306, 28, 116, 351, 310, 111], and we apply a reciprocal barrier function to its smoothed Lagrangian dual. For the error analysis and the numerical computations, we exploit the fact that the primal problem of the smoothed dual is a TRP.

The approximated DROP (1.1.4) is generally a nonsmooth, nonconvex optimization problem. Hence, algorithms for nonsmooth, nonconvex optimization can be applied to (1.1.4), such as, subgradient and bundle methods [175], gradient sampling algorithms [61], and quasi-Newton methods [216]. The approximated DROP (1.1.4) can also be reformulated as a nonlinear SDP (NSDP). We derive an equivalent reformulation of (1.1.4) as an NSDP in section 1.8.3 using Lagrangian duality for (1.1.5) and for (1.1.6) (see, e.g., [26, Chap. 4] and [53, sect. B.1]). We refer the reader to [352] for a survey on optimization methods for NSDPs.

We compare our algorithmic approach with the proximal bundle method `MPBNGC` [223, 224] applied to (1.1.4), and `PENLAB` [110] applied to an NSDP reformulation of (1.1.4) in section 1.6.

Related Work

A popular choice for constructing an ambiguity set is based on moment constraints of the parameters, such as the one in (1.1.2); see, e.g., [94, 295, 300, 344, 264, 283, 34]. Another approach is to define the set by probability measures close to a reference measure w.r.t. a certain distance [119, 293, 359, 340, 299], resulting in distance-based DROPs. Some nonconvex, data-driven, distance-based DROPs can equivalently be formulated as nonlinear programs with explicit objective and constraint functions, such as those considered in [340, sect. 2], [21] and [293, sect. 3.2]. We refer the reader to Shapiro [293] for an overview of distance-based DRO. A short discussion on the tractability of certain moment-based DROPs as opposed to DROPs defined by the Wasserstein distance is provided in [107, p. 117]. We refer the reader to [75, 94, 300, 344] for

further information on moment-based ambiguity sets, and to [340, pp. 243 and 249] and [119, p. 2] for discussions on the potential shortcomings of such sets.
Some specific classes of moment-based DROPs can be transformed into one-level problems using Lagrangian duality. For example, if ambiguity sets are conic representable, maximization problems w.r.t. probability measures become conic linear programs and, therefore, it can be transformed into minimization problems and concatenated with the upper-level problems [94, sect. 2.2]. If suitable assumptions, such as the convexity of the objective function w.r.t. design variables, are satisfied, the resulting optimization problem is tractable [94, 344, 346]. The reformulation of lower-level problems similar to those in (1.1.4) as linear matrix inequalities has been discussed in the supplementary material of [344].
Without the SDP (1.1.5) in (1.1.4), we obtain the robust optimization problem

$$\min_{x\in\mathbb{R}^n} a_j(x) + \psi_0(x) \quad \text{s.t.} \quad a_j(x) + \psi_j(x) \leq 0, \quad j \in J \setminus \{0\}. \tag{1.1.8}$$

Hence, our algorithmic approach can also be applied to (1.1.8). The robust nonlinear optimization problem (1.1.8) can be reformulated as an NSDP using either [23, Lem. 14.3.7] (see also [23, sect. 1.4]) or Proposition 1.8.3. Contributions on robust optimization may be divided into those exploiting concave dependence w.r.t. parameters (see, e.g., [22, 23, 26, 30]) and those developing schemes for robust nonlinear optimization (see, e.g., [96, 158, 358, 144, 218]).
We refer the reader to [218] for a recent survey on robust nonlinear optimization. Houska and Diehl [158] develop a numerical scheme for min-max optimization problems via sequential quadratic programming, and Ben-Tal and den Hertog [20] propose "sequential robust quadratic optimization". Robust nonlinear optimization without approximation techniques but heuristic numerical schemes are considered, for example, in [32, 33]. Some approaches are built on the methods of outer approximation [218, 232], originating from semiinfinite programming [305]. Derivative-free methods for robust nonlinear optimization are provided in [232, 83], a cutting plane method is proposed in [239], and a bundle method is developed in [199]. First-order Taylor expansions are used in [96, 358, 144] to obtain tractable approximations of the lower-level problems of robust nonlinear optimization problems. Instead of first-order expansions, second-order models are used, for example, in [298, 181, 209, 180]. These expansions may be more effective than first-order ones and may provide a trade-off between accuracy and tractability [4, 209]. The use of second-order expansions yields constraints such as those in (1.1.8), which are reformulated using its necessary and sufficient optimality conditions in [181, 209]. The resulting problem is a mathematical program with complementarity constraints and with linear matrix inequalities [181, 209]. These inequalities require the Hessian matrix of a Lagrangian function to be positive semidefinite which are reformulated using nonsmooth eigenvalue constraints in [181, 209]. We refer the reader to [171, 304] for an overview of numerical schemes for the solution of optimization problems with complementarity constraints.
The approximation of parameterized functions is a common approach to obtain simpler objective functions for optimization under uncertainty; see, e.g., [20, 96, 358, 181, 4] for robust optimization, [94, 31] for DRO, and [271] for reliability engineering.
Smoothing methods are popular schemes for nonconvex, nonsmooth optimization [58, 72, 349]. Our algorithmic scheme is related to those in [58, 59, 72, 349] in that we provide further examples of smoothing functions and apply their concepts and methodology. We use an NLP solver to compute approximate stationary points of a sequence of smoothed DROPs while the smoothing parameters converge to zero. Therefore, our algorithmic approach is similar to those in [72, 349]. Our scheme is built on the use of second-order approximations of the lower-level problems of the DROP (1.1.1). However, our approach allows the computation of stationary points of the approximated DROP (1.1.4) without the requirement that computationally available bounds on the Hessian matrix of $f_j(x,\cdot)$ are known as required by the approach developed by Houska and

Diehl [158]. Moreover, we do not require expensive numerical schemes as in [32, 33]. Our formulation avoids mathematical programs with complementarity constraints and with linear matrix inequalities as well as NSDPs, and the number of optimization variables in (1.1.4) is the same as in (1.1.1). Furthermore, we obtain smooth NLPs in standard form and their objective and constraint functions can efficiently be evaluated. Finally, different from the approach proposed in [158], our scheme allows the application of existing computer codes, making our approach applicable to many problems.

1.2 Smoothing Functions and Smoothing Method

We outline our algorithmic scheme to compute a first-order stationary point of (1.1.4). For each $j \in J$, we define $F_j : \mathbb{R}^n \to \mathbb{R}$ by

$$F_j(x) = a_j(x) + \varphi_j(x) + \psi_j(x), \tag{1.2.1}$$

where φ_j and ψ_j are defined in (1.1.6) and (1.1.5), respectively. Using the functions F_j, the approximated DROP (1.1.4) reads as

$$\min_{x\in\mathbb{R}^n} F_0(x) \quad \text{s.t.} \quad F_j(x) \le 0, \quad j \in J \setminus \{0\}, \tag{1.2.2}$$

which is generally a nonsmooth optimization problem. In the subsequent sections, we construct smooth approximations $\widetilde{F}_j : \mathbb{R}^n \times \mathbb{R}^3_{++} \to \mathbb{R}$ of F_j parameterized by $t \in \mathbb{R}^3_{++}$. The formal definition of the functions $\widetilde{F}_j$ is provided in (1.5.1). They are used in Algorithm 1 to compute a sequence of approximate KKT-points of the smoothed DROPs

$$\min_{x\in\mathbb{R}^n} \widetilde{F}_0(x,t) \quad \text{s.t.} \quad \widetilde{F}_j(x;t) \le 0, \quad j \in J \setminus \{0\}, \tag{1.2.3}$$

as $t \to 0^+$. Since these DROPs are smooth, we can apply state-of-the-art NLP software to compute KKT-tuples of them. Here, $(\bar{x}, \bar{\vartheta}) \in \mathbb{R}^n \times \mathbb{R}^{|J|-1}_{+}$ is a KKT-tuple of (1.2.2) if $\bar{\vartheta}_j F_j(\bar{x}) = 0$, $F_j(\bar{x}) \le 0$, $j \in J \setminus \{0\}$, and $0 \in \partial F_0(\bar{x}) + \sum_{j\in J\setminus\{0\}} \bar{\vartheta}_j \partial F_j(\bar{x})$. If a constraint qualification holds, these conditions are necessary optimality conditions for (1.2.2) [225, Cor. 5.1.8].
We construct smoothing functions of φ_j and of ψ_j, which satisfy the conditions of the following definition. Our notion of a smoothing function is based on those used in [73, Def. 3.1] and [72, Def. 1]; however, we allow for multiple smoothing parameters because the smoothing function for F_j constructed in section 1.5 depends on three.

Definition 1.2.1. *Let $\phi : \mathbb{R}^n \to \mathbb{R}$ be continuous. A function $\widetilde{\phi} : \mathbb{R}^n \times \mathbb{R}^m_{++} \to \mathbb{R}$ is a* smoothing function *for ϕ if, for each $t > 0$, $\widetilde{\phi}(\cdot; t)$ is continuously differentiable and there exists $\gamma : \mathbb{R}^m_+ \to \mathbb{R}_+$ with $\gamma(t) \to 0$ as $\mathbb{R}^m_{++} \ni t \to 0$ such that, for each $x \in \mathbb{R}^n$ and $t > 0$, we have $|\phi(x) - \widetilde{\phi}(x;t)| \le \gamma(t)$.*

Lemma 1.2.2. *If $\widetilde{\phi} : \mathbb{R}^n \times \mathbb{R}^m_{++} \to \mathbb{R}$ is a smoothing function for $\phi : \mathbb{R}^n \to \mathbb{R}$, then, for each $x \in \mathbb{R}^n$, $\lim_{\mathbb{R}^n \ni x^k \to x,\, t^k \to 0^+} \widetilde{\phi}(x^k; t^k) = \phi(x)$.*

Proof. Fix $(x^k) \subset \mathbb{R}^n$ and $(t^k) \subset \mathbb{R}^m_{++}$ with $x^k \to x$ as $k \to \infty$ and $t^k \to 0$ as $k \to \infty$, respectively. By assumption, ϕ is continuous, and there exists $\gamma : \mathbb{R}^m_+ \to \mathbb{R}_+$ such that $|\phi(x) - \widetilde{\phi}(x;t)| \le \gamma(t)$ for each $x \in \mathbb{R}^n$ and $t \in \mathbb{R}^m_{++}$. Combined with triangle inequality, we find that, for each $k \in \mathbb{N}$,

$$|\phi(x) - \phi(x^k; t^k)| \le |\phi(x^k) - \phi(x^k; t^k)| + |\phi(x^k) - \phi(x)| \le \gamma(t^k) + |\phi(x^k) - \phi(x)|.$$

Putting together the statements, we conclude that $\lim_{\mathbb{R}^n \ni x^k \to x,\, t^k \to 0^+} \widetilde{\phi}(x^k; t^k) = \phi(x)$. □

Algorithm 1 Smoothing method

Choose $t_0 \in \mathbb{R}^3_{++}$, $t_{\min} \in \mathbb{R}^3_+$, $\varepsilon_0 > 0$, $\varepsilon_{\min} \geq 0$ and $\rho \in (0,1)$.
For $k = 0, 1, \ldots$

1. Compute an ε_k-KKT-tuple (x^k, ϑ^k) of (1.2.3) for $t = t^k$.
2. If $t^k \leq t_{\min}$ and $\varepsilon_k \leq \varepsilon_{\min}$, then STOP and return (x^k, ϑ^k).
3. Compute $0 < t^{k+1} \leq \rho t^k$ and $\varepsilon_{k+1} = \rho \varepsilon_k$.

Lemma 1.2.2 shows that a smoothing function according to [73, Def. 3.1] is a smoothing function according to [72, Def. 1].
In Algorithm 1, it is sufficient to compute inexact KKT-tuples of (1.2.3), which may be important for an efficient numerical scheme for the approximated DROP (1.2.2). Different notions of approximate KKT-points have been proposed in the literature; see, e.g., [8, 103, 142]. We refer to $(x, \vartheta) \in \mathbb{R}^n \times \mathbb{R}^{|J|-1}$ as an ε-KKT-tuple of (1.2.3) if $\chi(x, \vartheta; t) \leq \varepsilon$. Here, the *criticality measure* $\chi : \mathbb{R}^n \times \mathbb{R}^{|J|-1} \times \mathbb{R}^3_{++} \to \mathbb{R}_+$ is defined by

$$\chi(x, \vartheta; t) = \max_{j \in J \setminus \{0\}} \left\{ \left\| \nabla_x \widetilde{F}_0(x; t) + \sum_{j \in J \setminus \{0\}} \vartheta_j \nabla_x \widetilde{F}_j(x; t) \right\|_\infty, \; |\min\{-\widetilde{F}_j(x; t), \vartheta_j\}| \right\}. \tag{1.2.4}$$

An important notion to establish convergence of Algorithm 1 to first-order stationary points of (1.2.2) is *gradient consistency*. Let $\widetilde{\phi} : \mathbb{R}^n \times \mathbb{R}^m_{++} \to \mathbb{R}$ be a smoothing function for the locally Lipschitz continuous function $\phi : \mathbb{R}^n \to \mathbb{R}$. Similar to the definition made by Chen [72, p. 73], we define

$$S_{\widetilde{\phi}}(x) = \operatorname{conv} \{ z \in \mathbb{R}^n : \; \exists \mathbb{R}^n \times \mathbb{R}^m_{++} \ni (x^k, t^k) \to (x, 0), \; \nabla_x \widetilde{\phi}(x^k; t_k) \to z \}. \tag{1.2.5}$$

Gradient consistency of $\widetilde{\phi}$ for ϕ requires the following relation to hold [58, 59, 72]:

$$S_{\widetilde{\phi}}(x) = \partial \phi(x) \quad \text{for all} \quad x \in \mathbb{R}^n. \tag{1.2.6}$$

The next lemma adapts [234, Lem. 3.2] to the notion of a smoothing function given in Definition 1.2.1. Lemma 1.2.3 is a consequence of [60, Lem. 3.1], whose proof is build on [276, Thm. 7.11 and Cor. 8.47].

Lemma 1.2.3. *If $\widetilde{\phi} : \mathbb{R}^n \times \mathbb{R}^m_{++} \to \mathbb{R}$ is a smoothing function for the locally Lipschitz continuous function $\phi : \mathbb{R}^n \to \mathbb{R}$, then, for all $x \in \mathbb{R}^n$, $\partial \phi(x) \subset S_{\widetilde{\phi}}(x)$.*

Proof. Fix $x \in \mathbb{R}^n$, and define $v = (1, \ldots, 1) \in \mathbb{R}^m$ and $\widetilde{g} : \mathbb{R}^n \times \mathbb{R}_{++} \to \mathbb{R}$ by $\widetilde{g}(x; t) = \widetilde{\phi}(x; tv)$. The function $\widetilde{g}$ is a smoothing function for ϕ. Combining Lemma 1.2.2 and [60, Lem. 3.1] yields $\partial \phi(x) \subset S_{\widetilde{g}}(x)$. Using (1.2.5), we obtain $S_{\widetilde{g}}(x) \subset S_{\widetilde{\phi}}(x)$. Hence $\partial \phi(x) \subset S_{\widetilde{\phi}}(x)$. □

In the next two sections, we construct smoothing functions for the optimal value functions defined in (1.1.6) and in (1.1.5). We show that these smoothing functions as well as their gradients can be evaluated efficiently, and establish their gradient consistency.

1.3 Smoothing Approach for the SDPs

We construct a smoothing function for the optimal value function φ_j defined in (1.1.5) exploiting the fact that the optimal values of the SDPs (1.1.5) can be computed analytically using the eigenvalues of a transformation of $C_j(x)$ for $x \in \mathbb{R}^n$ [350, Thm. 2.2]. We show that the smoothing function and its gradient can efficiently be evaluated, and that gradient consistency holds.

Proposition 1.3.1. *Let $C \in \mathbb{S}^p$ and suppose that $X_0,\ X_1 \in \mathbb{S}^p$ satisfy $X_0 \prec X_1$. Define $G = (X_1 - X_0)^{1/2} C (X_1 - X_0)^{1/2}$. Then*

$$C \bullet X_0 + \sum_{i=1}^{p} \min\{0, \lambda_i(G)\} = \min_{X \in \mathbb{S}^p} \{\, C \bullet X :\ X_0 \preccurlyeq X \preccurlyeq X_1 \,\}. \tag{1.3.1}$$

Moreover, $X^* = X_0 + (X_1 - X_0)^{1/2}[Q\mathrm{Diag}(y^*)Q^T](X_1 - X_0)^{1/2}$ is an optimal solution of (1.3.1), where $Q \in \mathbb{R}^{p\times p}$ fulfills $Q^TQ = I$ and $G = Q\mathrm{Diag}(\lambda(G))Q^T$, and $y^* \in \mathbb{R}^p$ satisfies $y_i^* = 1$ if $\lambda_i(G) < 0$ and $y_i^* = 0$ else.

Proof. The statements follow from an application of [350, Thm. 2.2]. □

The numerical results presented in section 1.8.4 indicate that the evaluation of the optimal value of (1.3.1) using the formula provided in (1.3.1) is significantly faster than some state-of-the-art SDP solvers. We define $G_j : \mathbb{R}^n \to \mathbb{S}^p$ by $G_j(x) = (\bar{\Sigma}_1 - \bar{\Sigma}_0)^{1/2} C_j(x) (\bar{\Sigma}_1 - \bar{\Sigma}_0)^{1/2}$. If $\bar{\Sigma}_0 \prec \bar{\Sigma}_1$, then (1.1.5) and (1.3.1) yield

$$\varphi_j(x) = (1/2)C_j(x) \bullet \bar{\Sigma}_0 + (1/2) \sum_{i=1}^{p} (\lambda_i(G_j(x)))_+ \quad \text{for all} \quad x \in \mathbb{R}^n. \tag{1.3.2}$$

In particular, φ_j is generally nonsmooth. We define $\widetilde{w} : \mathbb{R}^n \times \mathbb{R}_{++} \to \mathbb{R}$ by

$$\widetilde{w}(z;\tau) = \tau \sum_{i=1}^{p} \ln\left(1 + \exp\left(z_i/\tau\right)\right). \tag{1.3.3}$$

In order to prevent overflow when evaluating $\widetilde{w}(\cdot;\tau)$ and its gradient, we use, if $z > 0$, the identity $\tau \ln(1 + \exp(z/\tau)) = z + \tau \ln(\exp(-z/\tau) + 1)$, valid for each $z \in \mathbb{R}$ and $\tau > 0$.
Now, we show that $\widetilde{\varphi}_j : \mathbb{R}^n \times \mathbb{R}_{++} \to \mathbb{R}$ defined by

$$\widetilde{\varphi}_j(x;\tau) = (1/2)C_j(x) \bullet \bar{\Sigma}_0 + (1/2)\widetilde{w}(\lambda(G_j(x));\tau), \tag{1.3.4}$$

is a smoothing function for φ_j.

Theorem 1.3.2 ([234, Thm. 4.2])**.** *Let $\bar{\Sigma}_0 \prec \bar{\Sigma}_1$, $q \in \mathbb{N}$ and $j \in J$. Suppose that $C_j : \mathbb{R}^n \to \mathbb{S}^p$ is q-times continuously differentiable. Then the following conditions hold true:*

(a) *For each $(x, \tau) \in \mathbb{R}^n \times \mathbb{R}_{++}$, we have*

$$\varphi_j(x) \le \widetilde{\varphi}_j(x;\tau) \le \varphi_j(x) + (1/2)\tau p \ln 2, \tag{1.3.5}$$

where φ_j and $\widetilde{\varphi}_j$ are defined in (1.3.2) and (1.3.4), respectively.

(b) *The function $\widetilde{\varphi}_j$ is a smoothing function for φ_j, $\widetilde{\varphi}_j(\cdot;\tau)$ is q-times continuously differentiable for each $\tau > 0$, and gradient consistency holds.*

(c) *If $(x^k) \subset \mathbb{R}^n$ and $(\tau_k) \subset \mathbb{R}_{++}$ fulfill $x^k \to x \in \mathbb{R}^n$ and $\tau_k \to 0$ as $k \to \infty$, respectively, then there exists a convergent subsequence $(\nabla_x \widetilde{\varphi}_j(x^k;\tau^k))_K$ of $(\nabla_x \widetilde{\varphi}_j(x^k;\tau^k))$.*

Proof. (a) Since, for each $(z, \tau) \in \mathbb{R} \times \mathbb{R}_+$, we have $(z)_+ \le \tau \ln(1 + \exp(z/\tau)) \le (z)_+ + \tau \ln 2$ (see, e.g., [263, sect. 2]), we deduce (1.3.5) from (1.3.2).
(b) We establish that $\widetilde{\varphi}_j$ is a smoothing function for φ_j. Fix $\tau > 0$. The mapping λ is Lipschitz continuous [157, Cor. 6.3.8]. Using (1.3.2), we find that φ_j is the composition of locally Lipschitz continuous functions and, hence, it is locally Lipschitz continuous. Moreover, $\widetilde{w}(\cdot;\tau)$ defined in (1.3.3) is symmetric, and analytic because it is the composition of analytic functions.[1] Combined

[1] A function $h : \mathbb{R}^n \to \mathbb{R}$ is *symmetric* if its evaluations are invariant under coordinate permutations [217, p. 369].

with [319, Thm. 2.1], we find that $\widetilde{w}_\lambda(\cdot;\tau) = \widetilde{w}(\cdot;\tau) \circ \lambda$ is analytic. The chain rule implies that $\widetilde{\varphi}_j(\cdot;\tau) = (1/2)C_j(\cdot) \bullet \bar{\Sigma}_0 + (1/2)\widetilde{w}_\lambda(\cdot;\tau) \circ G_j$ is q-times continuously differentiable. Combined with the error estimate (1.3.5), we find that $\widetilde{\varphi}_j$ is a smoothing function for φ_j.
Now, we prove that gradient consistency holds, that is, we establish $S_{\widetilde{\varphi}_j}(x) = \partial\varphi_j(x)$ for each $x \in \mathbb{R}^n$, where $S_{\widetilde{\varphi}_j}(x)$ is defined in (1.2.5). Fix $x \in \mathbb{R}^n$. Since $\widetilde{\varphi}_j$ is locally Lipschitz continuous, Lemma 1.2.3 yields $\partial\varphi_j(x) \subset S_{\widetilde{\varphi}_j}(x)$.
It must yet be shown that $S_{\widetilde{\varphi}_j}(x) \subset \partial\varphi_j(x)$. We fix $z \in S_{\widetilde{\varphi}_j}(x)$. Then, by the definition of $S_{\widetilde{\varphi}_j}(x)$ provided in (1.2.5), there exist $(x^k) \subset \mathbb{R}^n$ and $(\tau_k) \subset \mathbb{R}_{++}$ converging to x and 0 as $k \to \infty$, respectively, and, moreover,

$$\nabla_x \widetilde{\varphi}_j(x^k;\tau_k) \to z \quad \text{as} \quad k \to \infty. \tag{1.3.6}$$

In order to show that $z \in \partial\varphi_j(x)$, we compute $\nabla_x \widetilde{\varphi}_j(x^k;\tau_k)$ for each $k \in \mathbb{N}_0$, and $\partial\varphi(x)$.
Fix $k \in \mathbb{N}_0$. The function $\widetilde{w}(\cdot;\tau_k)$ is continuously differentiable and symmetric and, hence, the classical chain rule and [214, Thm. 1.1] imply that the directional derivative $\mathrm{D}_x\widetilde{\varphi}_j(\cdot;\tau_k)h$ of $\widetilde{\varphi}_j(\cdot;\tau_k)$ w.r.t. x evaluated at x^k in direction $h \in \mathbb{R}^p$ is

$$\mathrm{D}_x\widetilde{\varphi}_j(x^k;\tau_k)h = (1/2)\bar{\Sigma}_0 \bullet \mathrm{D}C_j(x^k)h + (1/2)(Q_{j,k}M_{j,k}Q_{j,k}^T) \bullet \mathrm{D}G_j(x^k)h,$$

where $Q_{j,k} \in \mathbb{R}^{p\times p}$ fulfills $Q_{j,k}Q_{j,k}^T = I$ and $G_j(x^k) = Q_{j,k}\mathrm{Diag}(\lambda(G_j(x^k)))Q_{j,k}^T$, and where $M_{j,k} = \mathrm{Diag}(\nabla_x\widetilde{w}(\lambda(G_j(x^k));\tau_k))$. Using the adjoint operators $\mathrm{D}C_j(x^k)^*$ and $\mathrm{D}G_j(x^k)^*$ of $\mathrm{D}C_j(x^k)$ and $\mathrm{D}G_j(x^k)$, respectively, we obtain

$$\nabla_x\widetilde{\varphi}_j(x^k;\tau_k) = (1/2)\mathrm{D}C_j(x^k)^*\bar{\Sigma}_0 + (1/2)\mathrm{D}G_j(x^k)^*(Q_{j,k}M_{j,k}Q_{j,k}^T). \tag{1.3.7}$$

For every $P \in \mathbb{S}^p$, we have

$$\mathrm{D}C_j(x)^*P = \nabla_x(C_j(x) \bullet P) \quad \text{and} \quad \mathrm{D}G_j(x^k)^*P = \nabla_x(G_j(x^k) \bullet P). \tag{1.3.8}$$

Indeed, for each $s \in \mathbb{R}^n$ and $P \in \mathbb{S}^p$, we get

$$s^T\mathrm{D}C_j(x)^*P = P \bullet \mathrm{D}C_j(x)s = \mathrm{D}(C_j(x) \bullet P)s = s^T\nabla_x(C_j(x) \bullet P).$$

The second equation in (1.3.8) can be shown similarly.
Using (1.3.3), we obtain, for all $(z,\tau) \in \mathbb{R} \times \mathbb{R}_{++}$ and $i = 1,\ldots,p$,

$$(\nabla_z\widetilde{w}(z;\tau))_i = 1/(1 + \exp(-z_i/\tau)). \tag{1.3.9}$$

Hence $(\nabla_x\widetilde{w}(\lambda(G_j(x^k));\tau_k))$ is bounded. Moreover, $(Q_{j,k})$ is bounded. Combined with the (Lipschitz) continuity of λ [157, Cor. 6.3.8], we can assume without loss of generality that there exist $\bar{u}^j \in \mathbb{R}^p$ and $\bar{Q}_j \in \mathbb{R}^{p\times p}$ such that $\bar{Q}_j^T\bar{Q}_j = I$, $G_j(x) = \bar{Q}_j\mathrm{Diag}(\lambda(G_j(x)))\bar{Q}_j^T$, and

$$\nabla_x\widetilde{w}(\lambda(G_j(x^k));\tau_k) \to \bar{u}^j \quad \text{and} \quad Q_{j,k} \to \bar{Q}_j \quad \text{as} \quad k \to \infty.$$

In addition, for $i = 1,\ldots,p$, (1.3.9) implies that

$$(\nabla_x\widetilde{w}(\lambda(G_j(x^k));\tau_k))_i \to (\bar{u}^j)_i \in \begin{cases} \{0\} & \text{if} \quad \lambda_i(G_j(x)) < 0, \\ [0,1] & \text{if} \quad \lambda_i(G_j(x)) = 0, \\ \{1\} & \text{if} \quad \lambda_i(G_j(x)) > 0, \end{cases} \quad \text{as} \quad k \to \infty.$$

Combined with (1.3.7), (1.3.6), and the continuity of $\mathrm{D}C_j$ and $\mathrm{D}G_j$, we find that

$$\nabla_x\widetilde{\varphi}_j(x^k;\tau_k) \to (1/2)\mathrm{D}C_j(x)^*\bar{\Sigma}_0 + (1/2)\mathrm{D}G_j(x)^*\bar{Q}_j\mathrm{Diag}(\bar{u}^j)\bar{Q}_j^T = z \quad \text{as} \quad k \to \infty.$$

Next, we compute $\partial\varphi_j(x)$ using the representation of φ_j provided in (1.3.2). The function $\mathbb{S}^p \ni G \mapsto \sum_{i=1}^p (\lambda_i(G))_+$ is regular [215, Cor. 4], sums of regular functions are regular [79, Prop. 2.3.6], and continuously differentiable functions are regular [79, Prop. 2.3.6]. Combined with the chain rule [78, Thm. 2.3.10], and the formula for Clarke's subgradients of the eigenvalue mapping λ provided in [215, Thm. 8], we deduce

$$\partial\varphi_j(x) = \tfrac{1}{2}\big\{ \mathrm{D}C_j(x)^*\bar{\Sigma}_0 + \mathrm{D}G_j(x)^*Q\mathrm{Diag}(u)Q^T : Q \in O_j(x),\, u \in \partial w(\lambda(G_j(x))) \big\}, \quad (1.3.10)$$

where $O_j(x) = \{ Q \in \mathbb{R}^{p\times p} : Q^TQ = I,\, G_j(x) = Q\mathrm{Diag}(\lambda(G_j(x)))Q^T \}$ and $w : \mathbb{R}^p \to \mathbb{R}$ is defined by $w(z) = \sum_{i=1}^p (z)_+$. For each $z \in \mathbb{R}^p$, $i \in \{1,\ldots,p\}$, and every $g \in \partial w(z)$, we have $g_i = 0$ if $z_i < 0$, $g_i \in [0,1]$ if $z_i = 0$, and $g_i = 1$ if $z_i > 0$. Putting together the pieces, we find that $\bar{u}^j \in \partial w(\lambda(G_j(x)))$. Hence $z \in \partial\varphi_j(x)$.
(c) Under the stated assumptions, the proof of the second assertion can be used to deduce the existence of a subsequence of $(\nabla_x\widetilde{\varphi}_j(x^k;\tau^k))$. □

Based on a spectral decomposition of $G_j(x)$, the gradient $\nabla_x\widetilde{\varphi}_j(x;\tau)$ can efficiently be evaluated using (1.3.7) for $x \in \mathbb{R}^n$ and $\tau > 0$.

1.4 Smoothing Approach for the TRPs

We construct a smoothing function for the optimal value function $v : \mathbb{R}^n \to \mathbb{R}$ defined by

$$v(x) = \min_{s\in\mathbb{R}^p} \big\{ (1/2)s^TH(x)s + g(x)^Ts : \quad (1/2)\|s\|_2^2 \le (1/2)\Delta^2 \big\}, \quad (1.4.1)$$

where $g : \mathbb{R}^n \to \mathbb{R}^p$ and $H : \mathbb{R}^n \to \mathbb{S}^p$. Throughout, let $\Delta > 0$. The smoothing function for v allows us to construct one for the optimal value function defined in (1.1.6).
We obtain a smoothing function for (1.4.1) as the optimal value function of a "lifted" TRP, which results from a barrier formulation of a Lagrangian dual of (1.4.1). Since TRPs are theoretically and computationally tractable (see [28, sect. 2] and [238, sect. 5]), our construction implies that our smoothing function for v can be evaluated efficiently. Moreover, using Danskin's theorem, we show that the evaluations the gradient of this smoothing function is computationally tractable as well. In addition, we establish the gradient consistency for the smoothing function.

1.4.1 Lagrangian Dual of the TRP

We state the necessary and sufficient optimality conditions and review properties of the Lagrangian dual of the nominal TRP

$$\min_{s\in\mathbb{R}^p} (1/2)s^THs + g^Ts \quad \text{s.t.} \quad (1/2)\|s\|_2^2 \le (1/2)\Delta^2, \quad (1.4.2)$$

where $g = g(x_0) \in \mathbb{R}^p$, $H = H(x_0) \in \mathbb{S}^p$ for $x_0 \in \mathbb{R}^n$.

Theorem 1.4.1 ([301, Lems. 2.4 and 2.8])**.** *The point $s^* \in \mathbb{R}^p$ is an optimal solution to* (1.4.2) *if and only if there exists $\lambda^* \in \mathbb{R}$ such that*

$$(H + \lambda^*I)s^* = -g, \;\; \|s^*\|_2 \le \Delta, \;\; \lambda^*(\|s^*\|_2 - \Delta) = 0, \;\; \lambda^* \ge 0, \;\; H + \lambda^*I \succcurlyeq 0. \quad (1.4.3)$$

In addition, if (s^*,λ^*) fulfills (1.4.3) *and $\lambda^* > -\lambda_{\min}(H)$, then s^* is the unique optimal solution to* (1.4.2)*. Moreover, if (s_1^*,λ_1^*) and (s_2^*,λ_2^*) satisfy* (1.4.3)*, then $\lambda_1^* = \lambda_2^*$.*

The TRP (1.4.2) has an optimal solution because its feasible set is nonempty and compact, and its objective function is continuous. If (s^*,λ^*) satisfies (1.4.3), we refer to it as *optimal primal-dual solution* to (1.4.2).

Definition 1.4.2. *Let (s^*, λ^*) be an optimal primal-dual solution of* (1.4.2)*. If $\lambda^* = -\lambda_{\min}(H)$ holds, the* hard case *occurs for* (1.4.2)*, and otherwise the* easy case *occurs.*

The term "hard case" is due to Moré and Sorensen [238], and the term "easy case" has been used, for example, by Stern and Wolkowicz [306].
Now, we state a result on Lagrangian duality of (1.4.2).

Theorem 1.4.3 (see [310, Prop. 3.1, Thm. 3.3, and Cor. 3.4])**.** *A Lagrangian dual problem of* (1.4.2)*—phrased as a minimization problem—is*

$$\min_{\lambda \in \mathbb{R}} d(\lambda) \quad s.t. \quad H + \lambda I \succcurlyeq 0, \quad \lambda \geq 0, \tag{1.4.4}$$

where $d : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ is defined by

$$d(\lambda) = \begin{cases} \frac{1}{2} g^T (H + \lambda I)^+ g + \frac{1}{2}\Delta^2 \lambda & \text{if} \quad \lambda \geq (-\lambda_{\min}(H))_+,\ g \perp N(H + \lambda I), \\ \infty & \text{else.} \end{cases} \tag{1.4.5}$$

Moreover, (1.4.4) *has a unique optimal solution λ^*, which is the unique Lagrange multiplier corresponding to* (1.4.2)*. In addition, strong duality holds, that is, the optimal value of* (1.4.2) *equals $-d^*$, where d^* is the optimal value of* (1.4.4)*.*

We define the solution mapping $s : \mathbb{R} \to \mathbb{R}^p$ corresponding to (1.4.2) by

$$s(\lambda) = -(H + \lambda I)^+ g, \tag{1.4.6}$$

and state properties of the dual function d.

Lemma 1.4.4 ([234, Lem. 5.4])**.** *The following conditions hold true:*
(a) *The function d defined in* (1.4.5) *is convex, and $d(\lambda) \to \infty$ as $\lambda \to \infty$.*
(b) *If $\lambda > (-\lambda_{\min}(H))_+$, then d is twice continuously differentiable at λ, and*

$$d'(\lambda) = -(1/2)\|s(\lambda)\|_2^2 + (1/2)\Delta^2. \tag{1.4.7}$$

(c) *If $g \neq 0$, then $d''(\lambda) > 0$ for all $\lambda > (-\lambda_{\min}(H))_+$.*

Proof. The claims follow from [310, Prop. 3.2] and from the proof of [310, Thm. 3.3]. □

1.4.2 Barrier Formulation for the Dual of a TRP

We state a barrier problem for the dual (1.4.4) of the TRP (1.4.2) using a reciprocal barrier function. We show that an optimal solution to the barrier problem is an approximate solution to (1.4.4). In section 1.4.3, it is shown that the dual problem to the barrier problem corresponds to a "lifted" TRP. Since the dual to barrier problem is a TRP, it can be solved with any TRP solver, enabling us to define smoothing function for ψ_j (see (1.1.6)). The smoothing function and its gradient can efficiently be evaluated, allowing us to compute stationary points of the approximated DROP (1.2.2).
The barrier problem for (1.4.4) is

$$\min_{\lambda \in \mathbb{R}} d(\lambda) + \nu B_\eta(\lambda) \quad \text{s.t.} \quad \lambda > E(-H; \eta), \quad \lambda > 0, \tag{1.4.8}$$

where ν, $\eta > 0$, and the reciprocal barrier $B_\eta : ((E(-H;\eta))_+, \infty) \to \mathbb{R}$ is defined by

$$B_\eta(\lambda) = \frac{1}{\lambda} + \frac{1}{\lambda - E(-H;\eta)}. \tag{1.4.9}$$

Reciprocal barriers are also called inverse interior functions [112, sect. 3.1]. Here, $E : \mathbb{S}^p \times \mathbb{R}_{++} \to \mathbb{R}$ is an entropy function defined by

$$E(A;\eta) = \eta \ln \sum_{i=1}^{p} \exp(\lambda_i(A)/\eta). \tag{1.4.10}$$

The entropy function E has successfully been used in nonsmooth optimization [74, 247]. For each $\eta > 0$, $E(\cdot;\eta)$ is twice continuously differentiable [217, Thm. 4.2], and for each $A \in \mathbb{S}^p$,

$$\lambda_{\max}(A) \le E(A;\eta) \le \lambda_{\max}(A) + \eta \ln p; \tag{1.4.11}$$

[247, eqns. (17) and (18)]. The above properties, when combined with the (Lipschitz) continuity of $\lambda_{\max}$ [157, Cor. 6.3.8], imply that E is a smoothing function for $\lambda_{\max}$. In order to prevent overflow when evaluating $E(A;\eta)$, we use, if $\lambda_{\max}(A) > 0$, the identity $E(A;\eta) = \lambda_{\max}(A) + \eta \ln \sum_{i=1}^{p} \exp((\lambda_i(A) - \lambda_{\max}(A))/\eta)$, valid for each $A \in \mathbb{S}^p$ and $\eta > 0$.
We could use the self-concordant barrier function $((-\lambda_{\min}(H))_+, \infty) \ni \lambda \mapsto -\ln\lambda - \ln\det(H + \lambda I)$ in (1.4.8) which may not require the computation of $\lambda_{\min}(H)$ and to smooth $\lambda_{\min}$. However, the resulting primal problem would not be a TRP and requires, for example, an adapted version of the Moré–Sorensen algorithm [238, Alg. 3.2] for its numerical solution.
Before, we show that, for each ν, $\eta > 0$, the barrier problem (1.4.8) has a unique optimal solution, we observe that (1.4.11) yields, for each $A \in \mathbb{S}^p$ and $\eta > 0$,

$$\lambda_{\min}(A) = -\lambda_{\max}(-A) \ge -E(-A;\eta). \tag{1.4.12}$$

Lemma 1.4.5 ([234, Lem. 5.5])**.** *For each ν, $\eta > 0$, the barrier problem* (1.4.8) *has a unique optimal solution $\lambda^*(\nu,\eta)$, and $\lambda^*(\nu,\eta) > (E(-H;\eta))_+$, where E is defined in* (1.4.10)*.*

Proof. Fix ν, $\eta > 0$. We define the objective function of (1.4.8) by

$$B_{\nu,\eta} : ((E(-H;\eta))_+, \infty) \to \mathbb{R}, \quad B_{\nu,\eta} = d + \nu B_\eta, \tag{1.4.13}$$

where d and B_η are defined in (1.4.5) and (1.4.9), respectively. Fix $\lambda > (E(-H;\eta))_+$. Using (1.4.12), we have $(E(-H;\eta))_+ \ge (-\lambda_{\min}(H))_+$. Hence

$$B_{\nu,\eta}(\lambda) = \frac{1}{2} g^T (H + \lambda I)^{-1} g + \frac{1}{2}\Delta^2\lambda + \frac{\nu}{\lambda} + \frac{\nu}{\lambda - E(-H;\eta)} \ge \frac{1}{2}\Delta^2\lambda,$$

showing that $B_{\nu,\eta}(\lambda) \to \infty$ as $\lambda \to \infty$. From (1.4.5), (1.4.9), and (1.4.13), we obtain

$$B_{\nu,\eta}(\lambda) \ge \frac{\nu}{\lambda} + \frac{\nu}{\lambda - E(-H;\eta)} \to \infty \quad \text{as} \quad \lambda \to (E(-H;\eta))_+.$$

Consequently, (1.4.8) has an optimal solution $\lambda^*(\nu,\eta)$, and $\lambda^*(\nu,\eta) > (E(-H;\eta))_+$.
Now, we prove that $B_{\nu,\eta}$ defined in (1.4.13) is strictly convex. Lemma 1.4.4 implies that $B_{\nu,\eta}$ is twice continuously differentiable at λ with

$$\begin{aligned} B'_{\nu,\eta}(\lambda) &= -\frac{1}{2} g^T (H + \lambda I)^{-2} g - \frac{\nu}{\lambda^2} - \frac{\nu}{(\lambda - E(-H;\eta))^2} + \frac{1}{2}\Delta^2, \\ B''_{\nu,\eta}(\lambda) &= g^T (H + \lambda I)^{-3} g + \frac{2\nu}{\lambda^3} + \frac{2\nu}{(\lambda - E(-H;\eta))^3}. \end{aligned} \tag{1.4.14}$$

Hence $B''_{\nu,\eta}(\lambda) > 0$, which implies the strict convexity of $B_{\nu,\eta}$. Putting together the pieces, we conclude that $\lambda^*(\nu,\eta)$ is the unique optimal solution of (1.4.8). □

Theorem 1.4.6 ([234, Lem. 5.6]). *For fixed ν, $\eta > 0$, the following conditions hold true:*
(a) *We have*

$$\lambda^*(\nu,\eta) \geq \sqrt{2\nu}/\Delta \quad \text{and} \quad \lambda^*(\nu,\eta) - E(-H;\eta) \geq \sqrt{2\nu}/\Delta, \tag{1.4.15}$$

where $\lambda^*(\nu,\eta)$ is the optimal solution of (1.4.8) *and E is defined in* (1.4.10).
(b) *The point $\lambda^*(\nu,\eta)$ is a $(\sqrt{2\nu}\Delta + (1/2)\Delta^2\eta\ln p)$-optimal solution of* (1.4.4), *that is,*

$$d^* \leq d(\lambda^*(\nu,\eta)) \leq d^* + \sqrt{2\nu}\Delta + (1/2)\Delta^2\eta\ln p, \tag{1.4.16}$$

where d^* is the optimal value of (1.4.4), *and d is defined in* (1.4.5).
(c) *It holds that*

$$d^* \leq d(\lambda^*(\nu,\eta)) + \nu B_\eta(\lambda^*(\nu,\eta)) \leq d^* + 2\sqrt{2\nu}\Delta + (1/2)\Delta^2\eta\ln p, \tag{1.4.17}$$

where the barrier function B_η is defined in (1.4.9).

We apply the following result to prove Theorem 1.4.6.

Lemma 1.4.7 ([234, Lem. 5.7]). *Let η, $\epsilon > 0$ be arbitrary, and consider*

$$\min_{\lambda\in\mathbb{R}} d(\lambda) \quad s.t. \quad \lambda \geq \epsilon, \quad \lambda \geq E(-H;\eta) + \epsilon. \tag{1.4.18}$$

Then (1.4.18) *has a unique optimal solution $\bar\lambda_{\eta,\epsilon}$, and*

$$d^* \leq d(\bar\lambda_{\eta,\epsilon}) = d^*_{\eta,\epsilon} \leq d^* + (1/2)\Delta^2(\eta\ln p + \epsilon), \tag{1.4.19}$$

where d^* is the optimal value of (1.4.4) *and $d^*_{\eta,\epsilon}$ that of* (1.4.18).

Proof. We establish the existence and uniqueness of solutions to (1.4.18). If $g = 0$, then $d(\lambda) = \Delta^2\lambda/2$, and the optimal solution $\bar\lambda_{\eta,\epsilon}$ of (1.4.18) is $\bar\lambda_{\eta,\epsilon} = (E(-H;\eta))_+ + \epsilon$. If $g \neq 0$, then Lemma 1.4.4 and (1.4.12) imply that the objective of (1.4.18) is coercive, twice continuously differentiable on a neighborhood of the feasible set of (1.4.18), and $d''(\lambda) > 0$ for each $\lambda > (E(-H;\eta))_+$. Hence, there exists a unique optimal solution $\bar\lambda_{\eta,\epsilon}$ of (1.4.18).
Now, we establish (1.4.19). Since $\bar\lambda_{\eta,\epsilon} \geq (E(-H;\eta))_+ + \epsilon$, we have $d^* \leq d(\bar\lambda_{\eta,\epsilon})$. Moreover, if $\lambda^* > (E(-H;\eta))_+ + \epsilon$, then $d^* = d^*_{\eta,\epsilon}$, where λ^* is the optimal solution of (1.4.4). Hence, the remaining case to be considered is

$$(-\lambda_{\min}(H))_+ \leq \lambda^* \leq (E(-H;\eta))_+ + \epsilon.$$

We define $\bar\lambda = \lambda^* + \eta\ln p + \epsilon$. We have $\bar\lambda \geq \epsilon$. From (1.4.11), we deduce

$$E(-H;\eta) \leq -\lambda_{\min}(H) + \eta\ln p \leq \lambda^* + \eta\ln p$$

showing that $\bar\lambda \geq E(-H;\eta) + \epsilon$. Hence, $\bar\lambda$ is feasible for (1.4.18). Lemma 1.4.4 implies that d is convex, and differentiable at $\bar\lambda$. Hence

$$d(\lambda^*) - d(\bar\lambda) \geq d'(\bar\lambda)(\lambda^* - \bar\lambda) = -d'(\bar\lambda)(\eta\ln p + \epsilon)$$

resulting in

$$d(\lambda^*) + d'(\bar\lambda)(\eta\ln p + \epsilon) \geq d(\bar\lambda) \geq d(\bar\lambda_{\eta,\epsilon}).$$

Combined with (1.4.6), Lemma 1.4.4 and (1.4.7), we find that $d'(\bar\lambda) \leq (1/2)\Delta^2$ and, hence, (1.4.19) holds. □

In order to prove the estimates in (1.4.16), we use the fact that the functions $G_1 : (0, \infty) \to \mathbb{R}$ and $G_2 : (E(-H;\eta), \infty) \to \mathbb{R}$ defined by

$$G_1(\lambda) = -\ln \lambda, \quad \text{and} \quad G_2(\lambda) = -\ln(\lambda - E(-H;\eta))$$

are 1-self-concordant barrier functions of their domains [249, sect. 2.3.1, Ex. 2].

Proof of Theorem 1.4.6. (a) We establish (1.4.15). Recall that the objective of (1.4.8) is $B_{\nu,\eta}$, which is defined in (1.4.13). Lemma 1.4.5 implies that $B'_{\nu,\eta}(\lambda^*(\nu,\eta)) = 0$. Combined with (1.4.14), we have

$$g^T(H + \lambda^*(\nu,\eta)I)^{-2}g + \frac{2\nu}{\lambda^*(\nu,\eta)^2} + \frac{2\nu}{(\lambda^*(\nu,\eta) - E(-H;\eta))^2} = \Delta^2.$$

Lemma 1.4.5 and (1.4.12) further yield $H + \lambda^*(\nu,\eta)I \succ 0$. Hence

$$\frac{2\nu}{\lambda^*(\nu,\eta)^2} \le \Delta^2 \quad \text{and} \quad \frac{2\nu}{(\lambda^*(\nu,\eta) - E(-H;\eta))^2} \le \Delta^2,$$

showing the estimates in (1.4.15).
(b) We verify (1.4.16). Using (1.4.15), we find that $\lambda^*(\nu,\eta)$ is feasible for (1.4.4). Hence $d^* \le d(\lambda^*(\nu,\eta))$. Now, fix $\lambda > (E(-H;\eta))_+$. Both G_1 and G_2 defined prior the proof are 1-self-concordant for their domains. Combined with [249, Prop. 2.3.2], we find that

$$\begin{aligned} -\frac{1}{\lambda^*(\nu,\eta)}(\lambda - \lambda^*(\nu,\eta)) &= G_1'(\lambda^*(\nu,\eta))(\lambda - \lambda^*(\nu,\eta)) \le 1, \\ -\frac{1}{\lambda^*(\nu,\eta) - E(-H;\eta)}(\lambda - \lambda^*(\nu,\eta)) &= G_2'(\lambda^*(\nu,\eta))(\lambda - \lambda^*(\nu,\eta)) \le 1. \end{aligned} \tag{1.4.20}$$

Since $B'_{\nu,\eta}(\lambda^*(\nu,\eta)) = 0$ implies

$$d'(\lambda^*(\nu,\eta)) = -\nu B'_\eta(\lambda^*(\nu,\eta)),$$

the estimates in (1.4.15) and (1.4.20), and $\lambda^*(\nu,\eta) > (E(-H;\eta))_+$ ensure

$$\begin{aligned} d'(\lambda^*(\nu,\eta))(\lambda - \lambda^*(\nu,\eta)) &= -\nu B'_\eta(\lambda^*(\nu,\eta))(\lambda - \lambda^*(\nu,\eta)) \\ &= \frac{\nu}{\lambda^*(\nu,\eta)^2}(\lambda - \lambda^*(\nu,\eta)) + \frac{\nu}{(\lambda^*(\nu,\eta) - E(-H;\eta))^2}(\lambda - \lambda^*(\nu,\eta)) \\ &\ge -\frac{\nu}{\lambda^*(\nu,\eta)} - \frac{\nu}{\lambda^*(\nu,\eta) - E(-H;\eta)}. \end{aligned}$$

Combined with the convexity of d (see Lemma 1.4.4) and (1.4.15), we find that

$$\begin{aligned} d(\lambda^*(\nu,\eta)) - d(\lambda) &\le d'(\lambda^*(\nu,\eta))(\lambda^*(\nu,\eta) - \lambda) \\ &\le \frac{\nu}{\lambda^*(\nu,\eta)} + \frac{\nu}{\lambda^*(\nu,\eta) - E(-H;\eta)} \le \frac{2\nu}{\sqrt{2\nu}}\Delta = \sqrt{2\nu}\Delta. \end{aligned} \tag{1.4.21}$$

Now, we fix $\epsilon > 0$. Let $\bar{\lambda}_{\eta,\epsilon}$ the optimal solution of (1.4.18). Lemma 1.4.7 gives $\bar{\lambda}_{\eta,\epsilon} \ge (E(-H;\eta))_+ + \epsilon$. Furthermore, Lemma 1.4.7, (1.4.19) and (1.4.21) with $\lambda = \bar{\lambda}_{\eta,\epsilon}$ show that

$$d(\lambda^*(\nu,\eta)) \le d(\bar{\lambda}_{\eta,\epsilon}) + \sqrt{2\nu}\Delta \le d^* + \sqrt{2\nu}\Delta + (1/2)\Delta^2(\eta \ln p + \epsilon).$$

The latter inequalities hold for all $\epsilon > 0$ and, hence, we obtain (1.4.16).
(c) We show (1.4.17). Using (1.4.9) and (1.4.15), we have $\nu B_\eta(\lambda^*(\nu,\eta)) > 0$ and $\nu B_\eta(\lambda^*(\nu,\eta)) \le \sqrt{2\nu}\Delta$, and $\lambda^*(\nu,\eta)$ is feasible for (1.4.4). Hence (1.4.16) implies (1.4.17). □

The error estimates presented in Theorem 1.4.6 depend on $\ln p$ and on the prescribed trust-region radius Δ. Therefore, the data dependence is weak.

1.4.3 Smoothing Function for the TRPs

We show that the optimal value function $\widetilde{v} : \mathbb{R}^n \times \mathbb{R}^2_{++} \to \mathbb{R}$ defined by

$$\widetilde{v}(x;\nu,\eta) = \min_{\widetilde{s}\in\mathbb{R}^{p+2}} \left\{ (1/2)\widetilde{s}^T \widetilde{H}_\eta(x)\widetilde{s} + \widetilde{g}_\nu(x)^T\widetilde{s} : \ \ (1/2)\|\widetilde{s}\|_2^2 \le (1/2)\Delta^2 \right\}. \tag{1.4.22}$$

is a smoothing function for the function v in (1.4.1), and establish its gradient consistency. Here

$$\widetilde{H}_\eta(x) = \begin{bmatrix} H(x) & & \\ & 0 & \\ & & -E(-H(x);\eta) \end{bmatrix} \in \mathbb{S}^{p+2} \quad \text{and} \quad \widetilde{g}_\nu(x) = \begin{bmatrix} g(x) \\ \sqrt{2\nu} \\ \sqrt{2\nu} \end{bmatrix} \in \mathbb{R}^{p+2}, \tag{1.4.23}$$

and $E(\cdot;\eta)$ is defined in (1.4.10). Then, we apply these results to define a smoothing function for ψ_j defined in (1.1.6), to deduce its gradient consistency. In order to prove these properties, we use the fact that a Lagrangian dual of (1.4.22) is

$$\max_{\lambda\in\mathbb{R}} -d(\lambda;x) - \frac{\nu}{\lambda} - \frac{\nu}{\lambda - E(-H(x);\eta)} \quad \text{s.t.} \quad \lambda > (E(-H(x);\eta))_+, \tag{1.4.24}$$

where $x \in \mathbb{R}^n$ and $d : ((-E(H(x);\eta))_+, \infty) \times \mathbb{R}^n \to \mathbb{R}$ is defined by

$$d(\lambda;x) = (1/2)g(x)^T(H(x) + \lambda I)^{-1}g(x) + (1/2)\Delta^2\lambda. \tag{1.4.25}$$

Lemma 1.4.8 ([234, Lem. 5.8])**.** *Let $x \in \mathbb{R}^n$ and ν, $\eta > 0$ be arbitrary. Then the problem* (1.4.24) *has a unique optimal solution $\widetilde{\lambda}(x;\nu,\eta)$, and $\widetilde{\lambda}(x;\nu,\eta) > (E(-H(x);\eta))_+$. Moreover, the optimal value of* (1.4.22) *equals that of* (1.4.24)*, the easy case occurs for* (1.4.22)*, and*

$$\widetilde{v}(x;\nu,\eta) = -(1/2)\widetilde{g}_\nu(x)^T(\widetilde{H}_\eta(x) + \widetilde{\lambda}(x;\nu,\eta)I)^{-1}\widetilde{g}_\nu(x) - (1/2)\Delta^2\widetilde{\lambda}(x;\nu,\eta). \tag{1.4.26}$$

Proof. Lemma 1.4.5 implies that (1.4.24) has a unique minimizer $\widetilde{\lambda}(x;\nu,\eta)$, and $\widetilde{\lambda}(x;\nu,\eta) > (E(-H(x);\eta))_+$. From (1.4.12), we deduce $\lambda_{\min}(H(x)) \ge -E(-H(x);\eta)$, and (1.4.23) gives $\lambda_{\min}(\widetilde{H}_\eta(x)) = -(E(-H(x);\eta))_+$.
If $E(-H(x);\eta) > 0$, then (1.4.23) ensures $y = (0,\ldots,0,1) \in N(\widetilde{H}_\eta(x) - \lambda_{\min}(\widetilde{H}_\eta(x))I)$ and $y^T\widetilde{g}_\nu(x) \ne 0$. If $E(-H(x);\eta) \le 0$, then $w = (0,\ldots,0,1,0) \in N(\widetilde{H}_\eta(x) - \lambda_{\min}(\widetilde{H}_\eta(x))I)$ and $w^T\widetilde{g}_\nu(x) \ne 0$. Hence $\widetilde{g}_\nu(x) \not\perp N(\widetilde{H}_\eta(x) - \lambda_{\min}(\widetilde{H}_\eta(x))I)$. Theorem 1.4.3 yields $\widetilde{\lambda}(x;\nu,\eta) > (E(-H(x);\eta))_+$ and, hence, the easy case occurs for (1.4.22).
Next, for each $\lambda > (E(-H(x);\eta))_+$, (1.4.23) and (1.4.25) yield

$$d(\lambda;x) + \frac{\nu}{\lambda(x;\nu,\eta)} + \frac{\nu}{\lambda(x;\nu,\eta) - E(-H(x);\eta)} = \frac{1}{2}\widetilde{g}_\nu(x)^T(\widetilde{H}_\eta(x) + \lambda I)^{-1}\widetilde{g}_\nu(x) + \frac{1}{2}\Delta^2\lambda.$$

Combined with Theorem 1.4.3, we deduce the strong duality and (1.4.26). □

We establish an error estimate for $\widetilde{v}$ (see (1.4.22)), and show that it is a smoothing function for v (see (1.4.1)). We define, similarly to (1.4.6), the mapping $s : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$ by

$$s(\lambda;x) = -(H(x) + \lambda I)^+ g(x). \tag{1.4.27}$$

For ν, $\eta > 0$, let $(\widetilde{s}(x;\nu,\eta), \widetilde{\lambda}(x;\nu,\eta))$ be the optimal primal-dual solution of (1.4.22), where

$$\widetilde{\lambda}(\cdot;\nu,\eta) : \mathbb{R}^n \to \mathbb{R} \quad \text{and} \quad \widetilde{s}(\cdot;\nu,\eta) : \mathbb{R}^n \to \mathbb{R}^p. \tag{1.4.28}$$

From (1.4.3), Lemma 1.4.8, the block structure of $\widetilde{H}_\eta(x)$ ((1.4.23)), and (1.4.27), we obtain that, for all $x \in \mathbb{R}^n$,

$$\widetilde{s}(x;\nu,\eta) = (s(\widetilde{\lambda}(x;\nu,\eta);x), \widetilde{s}_{p+1}(x;\nu,\eta), \widetilde{s}_{p+2}(x;\nu,\eta)). \tag{1.4.29}$$

In particular, the first p components of $\widetilde{s}(x;\nu,\eta)$ are given by $s(\widetilde{\lambda}(x;\nu,\eta);x)$. Applying (1.4.3) and (1.4.23) yields

$$\widetilde{s}_{p+1}(x;\nu,\eta) = \frac{\sqrt{2\nu}}{\widetilde{\lambda}(x;\nu,\eta)} \quad \text{and} \quad \widetilde{s}_{p+2}(x;\nu,\eta) = \frac{\sqrt{2\nu}}{\widetilde{\lambda}(x;\nu,\eta) - E(-H(x);\eta)}. \tag{1.4.30}$$

Theorem 1.4.9 ([234, Thm. 5.9])**.** *Let ν, $\eta > 0$, and let $q \in \mathbb{N}$. Suppose that $g : \mathbb{R}^n \to \mathbb{R}^p$ and $H : \mathbb{R}^n \to \mathbb{S}^p$ are q-times continuously differentiable. Then the following conditions hold:*

(a) *For every $x \in \mathbb{R}^n$, we have*

$$v(x) \geq \widetilde{v}(x;\nu,\eta) \geq v(x) - 2\sqrt{2\nu}\Delta - (1/2)\Delta^2\eta \ln p, \tag{1.4.31}$$

where v is defined in (1.4.1) *and $\widetilde{v}$ in* (1.4.22)*.*

(b) *The mappings $\widetilde{s}(\cdot;\nu,\eta)$ and $\widetilde{\lambda}(\cdot;\nu,\eta)$ defined in* (1.4.28) *are $q-1$-times continuously differentiable, $\widetilde{v}(\cdot;\nu,\eta)$ is q-times continuously differentiable, and*

$$\nabla_x \widetilde{v}(x;\nu,\eta) = \nabla_x \wp(x,s)|_{s=s(\widetilde{\lambda};x)} + (1/2)(\widetilde{s}_{p+2})^2 \nabla_x(-E(-H(x);\eta)). \tag{1.4.32}$$

Here, $\wp : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}$ is defined by

$$\wp(x,s) = g(x)^T s + (1/2)s^T H(x)s, \tag{1.4.33}$$

and $(\widetilde{s},\widetilde{\lambda}) = (\widetilde{s}(x;\nu,\eta), \widetilde{\lambda}(x;\nu,\eta))$ is the optimal primal-dual solution of (1.4.22)*.*

(c) *The function $\widetilde{v}$ is a smoothing function for v.*

Proof. (a) Fix $x \in \mathbb{R}^n$. Combining Theorem 1.4.6 and Lemma 1.4.8, and (1.4.17) and (1.4.26) yields (1.4.31).

(b) Lemma 1.4.8 further implies $\widetilde{\lambda}(x;\nu,\eta) > (E(-H(x);\eta))_+$, implying that strict complementarity slackness holds for (1.4.22). Moreover, the entropy function $E(\cdot;\eta)$ defined in (1.4.10) is analytic since $z \mapsto \eta \ln \sum_{i=1}^p \exp(z_i/\eta)$ is analytic (see [319, Thm. 3.1]) and, therefore, the mapping $\widetilde{H}_\eta$ (see (1.4.23)) is q-times continuously differentiable. Hence, the implicit function theorem applies to the first-order optimality conditions (1.4.3) of (1.4.22) and implies that $\widetilde{\lambda}(\cdot;\nu,\eta)$ and $\widetilde{s}(\cdot;\nu,\eta)$ are $q-1$-times continuously differentiable.
Combined with (1.4.22), (1.4.23), (1.4.29), (1.4.33), and Danskin's theorem [46, Thm. 4.13 and Rem. 4.14] we find that $\widetilde{v}(\cdot;\nu,\eta)$ is differentiable and that its gradient is given by (1.4.32). Next, [154, Cor. 8.2] implies that $\widetilde{s}(\cdot;\nu,\eta)$ is continuous showing that $\nabla_x\widetilde{v}(\cdot;\nu,\eta)$ is continuous. Moreover, the chain rule and (1.4.22) imply that $\widetilde{v}(\cdot;\nu,\eta)$ is q-times continuously differentiable.

(c) The function v is continuous by [154, Thm. 7], $\widetilde{v}(\cdot;\nu,\eta)$ is continuously differentiable and, hence, (1.4.31) shows that $\widetilde{v}$ is a smoothing function for v. □

The next result asserts gradient consistency of the function $\widetilde{v}$ defined in (1.4.22).

Theorem 1.4.10 ([234, Thm. 5.10])**.** *If the hypotheses of Theorem 1.4.9 hold, then the following conditions are satisfied:*

(a) *Gradient consistency of $\widetilde{v}$ for v holds, where v is defined in* (1.4.1) *and $\widetilde{v}$ in* (1.4.22)*.*

(b) *If $(x^k) \subset \mathbb{R}^n$ fulfills $x^k \to x \in \mathbb{R}^n$, and (ν_k), $(\eta_k) \subset \mathbb{R}_{++}$ both converge to 0, then there exists a convergent subsequence $(\nabla_x\widetilde{v}(x^k;\nu_k,\eta_k))_K$ of $(\nabla_x\widetilde{v}(x^k;\nu_k,\eta_k))$.*

We apply Lemmas 1.4.11 and 1.4.12 to establish Theorem 1.4.10.

Lemma 1.4.11 ([234, Lem. 5.11])**.** *Let $(x^k) \subset \mathbb{R}^n$ and $(\eta_k) \subset \mathbb{R}_{++}$. Suppose that $x^k \to x \in \mathbb{R}^n$ as $k \to \infty$ and $\eta_k \to 0$ as $k \to \infty$. Moreover, let $A : \mathbb{R}^n \to \mathbb{S}^p$ be continuously differentiable. Then there exist a subsequence $(\nabla_x(E(\cdot;\eta_k) \circ A)(x^k))_K$ of $(\nabla_x(E(\cdot;\eta_k) \circ A)(x^k))$, $\theta_i \in [0,1]$ and $u_i \in \mathbb{R}^p$ such that*

$$\nabla_x(E(\cdot;\eta_k) \circ A)(x^k) \to \sum_{i=1}^{r} \theta_i \mathrm{D}A(x)^*[u_i u_i^T] \in \mathrm{D}A(x)^* \partial \lambda_{\max}(A(x)) \text{ as } K \ni k \to \infty.$$

Here, E is the entropy function defined in (1.4.10)*, $1 \le r \le r(A(x))$, $r(A(x))$ is the multiplicity of $\lambda_{\max}(A(x))$, $\sum_{i=1}^{r} \theta_i = 1$, and u_i are pairwise orthonormal eigenvectors of $A(x)$ corresponding to $\lambda_{\max}(A(x))$.*

Proof. We compute $\partial(\lambda_{\max} \circ A)(x)$ and $\nabla_x(E(\cdot;\eta_k) \circ A)(x^k)$. Since $\lambda_{\max}$ is convex and Lipschitz continuous, [79, Prop. 2.3.6] implies that $\lambda_{\max}$ is regular in the sense of [79, Def. 2.3.4]. Combined with the continuous differentiability of A, the chain rule [79, Thm. 2.3.9] yields

$$\partial(\lambda_{\max} \circ A)(x) = \mathrm{D}A(x)^* \partial \lambda_{\max}(A(x)). \tag{1.4.34}$$

Since $E(\cdot;\eta_k)$ is analytic [319, Thm. 3.1], the chain rule implies

$$\nabla_x(E(\cdot;\eta_k) \circ A)(x^k) = \mathrm{D}A(x^k)^* \nabla_A E(A(x^k);\eta_k). \tag{1.4.35}$$

We define $A_k = A(x^k)$ and $A = A(x)$, and we establish the existence of a subsequence $(\nabla_A E(A_k;\eta_k))_K$ of $(\nabla_A E(A_k;\eta_k))$ such that

$$\nabla_A E(A_k;\eta_k) \to \sum_{i=1}^{r} \theta_i u_i u_i^T \in \partial \lambda_{\max}(A) \quad \text{as} \quad K \ni k \to \infty. \tag{1.4.36}$$

For all $k \in \mathbb{N}_0$, we have

$$\nabla_A E(A_k;\eta_k) = \sum_{i=1}^{p} \theta_{i,k} u_i(A_k) u_i(A_k)^T, \quad \text{and} \quad \theta_{i,k} = \frac{\exp \frac{\lambda_i(A_k) - \lambda_{\max}(A_k)}{\eta_k}}{\sum_{i=1}^{p} \exp \frac{\lambda_i(A_k) - \lambda_{\max}(A_k)}{\eta_k}},$$

where $A_k u_i(A_k) = \lambda_{\max}(A_k) u_i(A_k)$, and $u_i(A_k)$ are pairwise orthonormal for $i = 1, \ldots, p$ [247, sect. 4]. We have $\sum_{i=1}^{p} \theta_{i,k} = 1$ and $\theta_{i,k} \in [0,1]$. Hence, we can assume without loss of generality that for $i \in \{1, \ldots, p\}$, it holds that $u_i(A_k) \to u_i \in \mathbb{R}^p$ as $k \to \infty$, and $\theta_{i,k} \to \theta_i \in [0,1]$ as $k \to \infty$, $\|u_i\|_2 = 1$, and $\sum_{i=1}^{p} \theta_i = 1$. Combined with $A_k u_i(A_k) = \lambda_i(A_k) u_i(A_k)$, valid for each $k \in \mathbb{N}_0$, $A_k \to A$ as $k \to \infty$, and the (Lipschitz) continuity of λ [157, Cor. 6.3.8], we find that u_i is an eigenvector of A corresponding to $\lambda_i(A)$. Moreover $0 = u_i(A_k)^T u_j(A_k) \to u_i^T u_j$ as $k \to \infty$ for each $i \neq j$ implies that u_i are pairwise orthogonal. Now, let $i \in \{1, \ldots, p\}$ be an index such that $\lambda_i(A) < \lambda_{\max}(A)$, that is, $i > r(A)$. We obtain that $\lambda_i(A_k) - \lambda_{\max}(A_k) \le (\lambda_i(A) - \lambda_{\max}(A))/2 < 0$ for all sufficiently large $k \in \mathbb{N}_0$. Therefore $\theta_{i,k} \to 0$ as $k \to \infty$ implying $\theta_i = 0$. Combined with

$$\partial \lambda_{\max}(A) = \operatorname{conv} \left\{ uu^T : \; Au = \lambda_{\max}(A) u, \; \|u\|_2 = 1, \; u \in \mathbb{R}^p \right\}$$

(see [247, sect. 4]), we obtain (1.4.36). We have $\mathrm{D}A(x^k) \to \mathrm{D}A(x)$ as $k \to \infty$ and, hence, (1.4.34), (1.4.35), and (1.4.36) imply the claim. □

Lemma 1.4.12 ([234, Lem. 5.12])**.** *Let the hypotheses of Theorem 1.4.9 hold. Let $(x^k) \subset \mathbb{R}^n$ and (ν_k), $(\eta_k) \subset \mathbb{R}_{++}$. Suppose that $x^k \to \bar{x} \in \mathbb{R}^n$ and ν_k, $\eta_k \to 0$ as $k \to \infty$. We define $(\widetilde{s}^k, \widetilde{\lambda}_k) = (\widetilde{s}(x^k;\nu_k,\eta_k), \widetilde{\lambda}(x^k;\nu_k,\eta_k))$, where $(\widetilde{s}(x;\nu,\eta), \widetilde{\lambda}(x;\nu,\eta))$ is defined in* (1.4.28)*. Then the following conditions hold true.*

(a) *The sequence* $(\widetilde{s}^k, \widetilde{\lambda}_k)_{\mathbb{N}_0}$ *has a convergent subsequence* $(\widetilde{s}^k, \widetilde{\lambda}_k)_K$. *In particular, there exist* $(\bar{s}, \bar{\lambda}) \in \mathbb{R}^p \times \mathbb{R}_+$ *and* $\bar{\alpha}$, $\bar{\beta} \in \mathbb{R}$ *such that*

$$\widetilde{s}^k = (s(\widetilde{\lambda}_k; x^k), \widetilde{s}^k_{p+1}, \widetilde{s}^k_{p+2}) \to (\bar{s}, \bar{\beta}, \bar{\alpha}) \quad \text{and} \quad \widetilde{\lambda}_k \to \bar{\lambda} \quad \text{as } K \ni k \to \infty. \tag{1.4.37}$$

(b) *If* $\bar{\lambda} > -\lambda_{\min}(H(\bar{x}))$, *then* $\bar{\alpha} = 0$ *and* $(\bar{s}, \bar{\lambda})$ *is an optimal primal-dual solution of* (1.4.1) *for* $x = \bar{x}$.

(c) *If* $\bar{\lambda} = -\lambda_{\min}(H(\bar{x}))$ *and* $w_i \in \mathbb{R}^p$ *for* $i = 1, \dots, r \in \mathbb{N}$, *are pairwise orthonormal eigenvectors of* $H(\bar{x})$ *corresponding to* $\lambda_{\min}(H(\bar{x}))$, *then* $(\bar{s} + \gamma_i^+ w_i, \bar{\lambda})$ *and* $(\bar{s} + \gamma_i^- w_i, \bar{\lambda})$ *are optimal primal-dual solutions of* (1.4.1) *for* $x = \bar{x}$, *where*

$$\gamma_i^+ = -w_i^T \bar{s} + ((w_i^T \bar{s})^2 + \bar{\alpha}^2)^{1/2} \quad \text{and} \quad \gamma_i^- = -w_i^T \bar{s} - ((w_i^T \bar{s})^2 + \bar{\alpha}^2)^{1/2}. \tag{1.4.38}$$

Proof. (a) We show that $(\widetilde{s}^k, \widetilde{\lambda}_k)_{\mathbb{N}_0}$ is bounded. Since $\|\widetilde{s}^k\|_2 \le \Delta$, $(\widetilde{s}^k)$ is bounded. Now, fix $k \in \mathbb{N}_0$. Lemma 1.4.8 gives $\widetilde{\lambda}_k = \widetilde{\lambda}(x^k; \nu_k, \eta_k) > (E(-H(x^k); \eta))_+$ and, hence, (1.4.12) implies

$$\widetilde{\lambda}_k > -(\lambda_{\min}(H(x^k)))_+. \tag{1.4.39}$$

Combined with (1.4.26), Lemma 1.4.8, and (1.4.39), we obtain $\widetilde{v}(x^k; \nu_k, \eta_k) \le -(1/2)\Delta^2 \widetilde{\lambda}_k \le 0$. Theorem 1.4.9 yields $\widetilde{v}(x^k; \nu_k, \eta_k) \to v(\bar{x})$ as $k \to \infty$. Combined with $\Delta > 0$, we find that $(\widetilde{\lambda}_k)$ is bounded. In particular, $(\widetilde{s}^k, \widetilde{\lambda}_k)_{\mathbb{N}_0}$ is bounded and has a convergent subsequence. Hence, (1.4.29) implies (1.4.37) for some $(\bar{s}, \bar{\lambda}) \in \mathbb{R}^p \times \mathbb{R}_+$ and $\bar{\alpha}$, $\bar{\beta} \in \mathbb{R}$.
Next, we prepare the proofs of the second and third assertion. Using (1.4.14) and (1.4.30), we find that the first-order necessary optimality condition of (1.4.22) is

$$\Delta^2 = \|s(\widetilde{\lambda}_k; x^k)\|_2^2 + \frac{2\nu_k}{\widetilde{\lambda}_k^2} + \frac{2\nu_k}{(\widetilde{\lambda}_k - E(-H(x^k); \eta_k))^2} = \|\widetilde{s}^k\|_2^2.$$

Therefore, (1.4.37) ensures

$$\Delta^2 = \|\widetilde{s}^k\|_2^2 \to \|\bar{s}\|_2^2 + \bar{\beta}^2 + \bar{\alpha}^2 \quad \text{as} \quad K \ni k \to \infty, \tag{1.4.40}$$

and (1.4.39) implies

$$H(\bar{x}) + \bar{\lambda} I \succcurlyeq 0 \quad \text{and} \quad \bar{\lambda} \ge 0. \tag{1.4.41}$$

Moreover, using (1.4.27) and (1.4.39), we have

$$0 = (H(x^k) + \widetilde{\lambda}_k I) s(\widetilde{\lambda}_k; x^k) + g(x^k) \to (H(\bar{x}) + \bar{\lambda} I)\bar{s} + g(\bar{x}) \text{ as } K \ni k \to \infty. \tag{1.4.42}$$

(b) We verify that $(\bar{s}, \bar{\lambda})$ is an optimal primal-dual solution of (1.4.1) for $x = \bar{x}$ and $\bar{\alpha} = 0$ if $\bar{\lambda} > -\lambda_{\min}(H(\bar{x}))$. By assumption $H(\bar{x}) + \bar{\lambda} I$ is invertible and, hence, (1.4.42) implies that $\bar{s}$ is the unique solution to $(H(\bar{x}) + \bar{\lambda} I)\bar{s} = -g(\bar{x})$. Therefore, (1.4.27) and (1.4.42) result in $s(\bar{\lambda}; \bar{x}) = \bar{s}$. Moreover, (1.4.40) implies that $\|\bar{s}\|_2 \le \Delta$.
If $\bar{\lambda} > 0$, then the continuity of $\lambda_{\min}$, $\bar{\lambda} > -\lambda_{\min}(H(\bar{x}))$, $\widetilde{\lambda}_k \to \bar{\lambda}$ as $K \ni k \to \infty$ and (1.4.11) imply that $\widetilde{\lambda}_k \ge \bar{\lambda}/2 > 0$ and $\widetilde{\lambda}_k - E(-H(x^k); \eta_k) \ge (\bar{\lambda} + \lambda_{\min}(H(\bar{x})))/2 > 0$ for all sufficiently large $k \in K$. Combined with (1.4.30), we find that

$$\widetilde{s}^k_{p+1} = \frac{\sqrt{2\nu_k}}{\widetilde{\lambda}_k} \to 0 \text{ and } \widetilde{s}^k_{p+2} = \frac{\sqrt{2\nu_k}}{\widetilde{\lambda}_k - E(-H(x^k); \eta_k)} \to 0 \text{ as } K \ni k \to \infty, \tag{1.4.43}$$

and, therefore, $\bar{\alpha}$, $\bar{\beta} = 0$. Now, (1.4.40) yields $\Delta^2 = \|\bar{s}\|_2^2$. Hence, $(s(\bar{\lambda}; \bar{x}), \bar{\lambda})$ satisfies $\bar{\lambda}(\|\bar{s}\|_2^2 - \Delta^2) = 0$ and, therefore, it fulfills (1.4.3) implying that it is an optimal primal-dual solution of (1.4.1) for $x = \bar{x}$ by Theorem 1.4.1.

(c) We establish that $(\bar{s}+\gamma_i^+ w_i, \bar{\lambda})$ and $(\bar{s}+\gamma_i^- w_i, \bar{\lambda})$ are optimal primal-dual solutions of (1.4.1) for $x = \bar{x}$ and $i = 1, \ldots, r$ if $\bar{\lambda} = -\lambda_{\min}(H(\bar{x}))$. Fix $i \in \{1, \ldots, r\}$. The numbers γ_i^+ and γ_i^- defined in (1.4.38) solve

$$\gamma_i^2 + 2\gamma_i w_i^T \bar{s} - \bar{\alpha}^2 = 0.$$

Combined with $\|w_i\|_2 = 1$ and (1.4.40), we obtain, for $\gamma_i \in \{\gamma_i^-, \gamma_i^+\}$,

$$\|\bar{s} + \gamma_i w_i\|_2^2 = \|\bar{s}\|_2^2 + 2\gamma_i w_i^T \bar{s} + \gamma_i^2 = \Delta^2 - \bar{\alpha}^2 - \bar{\beta}^2 + 2\gamma_i w_i^T \bar{s} + \gamma_i^2 \leq \Delta^2 \tag{1.4.44}$$

with equality if $\bar{\beta} = 0$. Moreover, (1.4.42) and $(H(\bar{x}) + \bar{\lambda} I) w_i = 0$ result in

$$(H(\bar{x}) + \bar{\lambda} I)(\bar{s} + \gamma_i w_i) = (H(\bar{x}) + \bar{\lambda} I)\bar{s} = -g(\bar{x}). \tag{1.4.45}$$

If $\bar{\lambda} > 0$, then (1.4.43) and (1.4.37) imply $\bar{\beta} = 0$. Hence, (1.4.44) ensures $\bar{\lambda}(\|\bar{s}+\gamma_i w_i\|_2^2 - \Delta^2) = 0$. Combined with (1.4.41), (1.4.42), (1.4.44), (1.4.45), and Theorem 1.4.1, we conclude that $(\bar{s} + \gamma_i^+ w_i, \bar{\lambda})$ and $(\bar{s} + \gamma_i^- w_i, \bar{\lambda})$ are optimal primal-dual solutions of (1.4.1) for $x = \bar{x}$. □

The proof of Theorem 1.4.9 requires the gradient of the function $\wp$ defined in (1.4.33). Using (1.4.46) and a similar derivation as in (1.3.8), we obtain

$$\nabla_x \wp(x, s) = \nabla_x g(x)^T s + \tfrac{1}{2} \nabla_x s^T H(x) s = \nabla_x g(x)^T s + \tfrac{1}{2} \mathrm{D}H(x)^*[ss^T]. \tag{1.4.46}$$

Proof of Theorem 1.4.10. (a) Fix $\bar{x} \in \mathbb{R}^n$. The optimal value function v defined in (1.4.1) is locally Lipschitz continuous [109, Thm. 4.1], and, hence, $\partial v(\bar{x})$ is well-defined [225, Def. 3.1.3]. Combined with the definition of $\wp$ provided in (1.4.33), and [78, Thm. 2.1], we obtain

$$\partial v(\bar{x}) = \mathrm{conv}\,\{\, \nabla_x \wp(\bar{x}, s^*) : \, s^* \in \mathrm{Sol}_{\mathrm{TR}}^*(\bar{x}) \,\}, \tag{1.4.47}$$

where $\mathrm{Sol}_{\mathrm{TR}}^*(\bar{x})$ is the set of optimal solutions of the parameterized TRP (1.4.1) for $x = \bar{x}$.
Next, we establish gradient consistency, that is, we prove $\partial v(\bar{x}) = S_{\widetilde{v}}(\bar{x})$. Here, $S_{\widetilde{v}}(x)$ is defined in (1.2.5). Since v is locally Lipschitz continuous, Lemma 1.2.3 gives $\partial v(\bar{x}) \subset S_{\widetilde{v}}(\bar{x})$. It must yet be shown that $S_{\widetilde{v}}(\bar{x}) \subset \partial v(\bar{x})$, which we establish by distinguishing whether the easy or the hard case occurs for (1.4.1) with $x = \bar{x}$.
Fix $z \in S_{\widetilde{v}}(\bar{x})$. By assumption, there exist $(x^k) \subset \mathbb{R}^n$ and (ν_k), $(\eta_k) \subset \mathbb{R}_{++}$ with $x^k \to \bar{x}$ and ν_k, $\eta_k \to 0$, and

$$\nabla_x \widetilde{v}(x^k; \nu_k, \eta_k) \to z \quad \text{as} \quad k \to \infty. \tag{1.4.48}$$

Next, we derive a formula for z that allows us to deduce $z \in \partial v(\bar{x})$.
Lemma 1.4.12 implies that the sequence $(\widetilde{s}^k, \widetilde{\lambda}_k)_{\mathbb{N}_0}$ of optimal primal-dual solutions $(\widetilde{s}^k, \widetilde{\lambda}_k)$ of (1.4.22) for $(x, \nu, \eta) = (x^k, \nu_k, \eta_k)$ has a convergent subsequence $(\widetilde{s}^k, \widetilde{\lambda}_k)_K$, and $(s(\widetilde{\lambda}_k; x^k), \widetilde{\lambda}_k)_K$ converges to $(\bar{s}, \bar{\lambda})$ and $\widetilde{s}_{p+2} \to \bar{\alpha}$ as $K \ni k \to \infty$, where $\bar{s} \in \mathbb{R}^p$, $\bar{\lambda} \geq 0$ and $\bar{\alpha} \in \mathbb{R}$, and $s(\lambda; x)$ is defined in (1.4.27). In addition, Lemma 1.4.11 applies to $A = -H$ and yields the existence of a subsequence $(\nabla_x(E(-H(x^k); \eta_k)))_{K'}$ of $(\nabla_x(E(-H(x^k); \eta_k)))_K$ such that

$$\nabla_x(E(-H(x^k); \eta_k)) \to -\sum_{i=1}^r \theta_i \mathrm{D}H(\bar{x})^*[w_i w_i^T] \quad \text{as} \quad K' \ni k \to \infty, \tag{1.4.49}$$

where $1 \leq r \leq r(A(\bar{x}))$, $r(A(\bar{x}))$ is the multiplicity of $\lambda_{\max}(A(\bar{x}))$, $\theta_i \in [0, 1]$, and $\sum_{i=1}^r \theta_i = 1$. Moreover, w_i are pairwise orthonormal eigenvectors of $A(\bar{x}) = -H(\bar{x})$ corresponding to $\lambda_{\max}(A(\bar{x})) = -\lambda_{\min}(H(\bar{x}))$. Combined with (1.4.32), (1.4.48), (1.4.49), and the fact that g and H are continuously differentiable, we find that

$$\nabla_x \widetilde{v}(x^k; \nu_k, \eta_k) \to \nabla_x \wp(\bar{x}, \bar{s}) + (\bar{\alpha}^2/2) \sum_{i=1}^r \theta_i \mathrm{D}H(\bar{x})^*[w_i w_i^T] = z \text{ as } K' \ni k \to \infty. \tag{1.4.50}$$

If the easy case occurs for (1.4.1) with $x = \bar{x}$, then Lemma 1.4.12 implies that $\bar{s} \in \mathrm{Sol}^*_{\mathrm{TR}}(\bar{x})$ and $\bar{\alpha} = 0$. By applying (1.4.47), (1.4.48) and (1.4.50), we find that $z \in \partial v(\bar{x})$.
If the hard case occurs for (1.4.1), then Lemma 1.4.12 further implies that $\bar{s}+\gamma_i^+ w_i$ and $\bar{s}+\gamma_i^- w_i$ are optimal solutions of (1.4.1) for $x = \bar{x}$, where γ_i^+ and γ_i^- are defined in (1.4.38). If $\bar{\alpha} = 0$, then (1.4.38) implies that either γ_i^+ or γ_i^- is zero and, hence, $\bar{s}$ is an optimal solution of (1.4.1) for $x = \bar{x}$. Consequently, (1.4.47), (1.4.48) and (1.4.50) imply that $z \in \partial v(\bar{x})$. If $\bar{\alpha} > 0$, then (1.4.38) yields $\gamma_i^+ - \gamma_i^- = 2((w_i^T \bar{s})^2 + \bar{\alpha}^2)^{1/2} > 0$. We define

$$\tau_i^+ = \frac{-\gamma_i^-}{\gamma_i^+ - \gamma_i^-} \quad \text{and} \quad \tau_i^- = \frac{\gamma_i^+}{\gamma_i^+ - \gamma_i^-}. \tag{1.4.51}$$

Owing to (1.4.38), we have $\gamma_i^+ > 0$ and $\gamma_i^- < 0$. Together with (1.4.51), we obtain

$$\begin{aligned} &\tau_i^+ > 0, \quad \tau_i^- > 0, \quad \tau_i^+ + \tau_i^- = 1, \\ &\tau_i^+ \gamma_i^+ + \tau_i^- \gamma_i^- = \frac{-\gamma_i^- \gamma_i^+ + \gamma_i^+ \gamma_i^-}{\gamma_i^+ - \gamma_i^-} = 0, \quad \text{and} \quad \tau_i^+ (\gamma_i^+)^2 + \tau_i^- (\gamma_i^-)^2 = \bar{\alpha}^2. \end{aligned} \tag{1.4.52}$$

Using (1.4.33) and (1.4.46), we obtain, for $\gamma_i \in \{\gamma_i^-, \gamma_i^+\}$,

$$\begin{aligned} \nabla_x \wp(\bar{x}, \bar{s} + \gamma_i w_i) = {} & \nabla_x g(\bar{x})^T \bar{s} + (1/2)\mathrm{D}H(\bar{x})^*[\bar{s}\bar{s}^T] + \gamma_i \nabla_x g(\bar{x})^T w_i \\ & + (1/2)\gamma_i \mathrm{D}H(\bar{x})^*[w_i \bar{s}^T + \bar{s} w_i^T] + (1/2)(\gamma_i)^2 \mathrm{D}H(\bar{x})^*[w_i w_i^T] \end{aligned}$$

resulting in

$$\begin{aligned} &\tau_i^+ \nabla_x \wp(\bar{x}, \bar{s} + \gamma_i^+ w_i) + \tau_i^- \nabla_x \wp(\bar{x}, \bar{s} + \gamma_i^- w_i) \\ &\quad = (\tau_i^- + \tau_i^+) \nabla_x g(\bar{x})^T \bar{s} + (1/2)(\tau_i^- + \tau_i^+) \mathrm{D}H(\bar{x})^*[\bar{s}\bar{s}^T] \\ &\qquad + (\tau_i^+ \gamma_i^+ + \tau_i^- \gamma_i^-) \nabla g(\bar{x})^T w_i + (1/2)(\tau_i^+ \gamma_i^+ + \tau_i^- \gamma_i^-) \mathrm{D}H(\bar{x})^*[w_i \bar{s}^T + \bar{s} w_i^T] \\ &\qquad + (1/2)(\tau_i^+ (\gamma_i^+)^2 + \tau_i^- (\gamma_i^-)^2) \mathrm{D}H(\bar{x})^*[w_i w_i^T]. \end{aligned}$$

Combined with (1.4.52), we get

$$\begin{aligned} &\tau_i^+ \nabla_x \wp(\bar{x}, \bar{s} + \gamma_i^+ w_i) + \tau_i^- \nabla_x \wp(\bar{x}, \bar{s} + \gamma_i^- w_i) \\ &\quad = \nabla_x g(\bar{x})^T \bar{s} + (1/2)\mathrm{D}H(\bar{x})^*[\bar{s}\bar{s}^T] + (\bar{\alpha}^2/2)\mathrm{D}H(\bar{x})^*[w_i w_i^T], \end{aligned}$$

implying, with $\sum_{i=1}^r \theta_i = 1$ and (1.4.46), that

$$\begin{aligned} &\sum_{i=1}^r \theta_i \tau_i^+ \nabla_x \wp(\bar{x}, \bar{s} + \gamma_i^+ w_i) + \sum_{i=1}^r \theta_i \tau_i^- \nabla_x \wp(\bar{x}, \bar{s} + \gamma_i^- w_i) \\ &\qquad = \nabla_x \wp(\bar{x}, \bar{s}) + (\bar{\alpha}^2/2) \sum_{i=1}^r \theta_i \mathrm{D}H(\bar{x})^*[w_i w_i^T]. \end{aligned} \tag{1.4.53}$$

Moreover, using (1.4.52), we have $\sum_{i=1}^r \theta_i \tau_i^+ + \sum_{i=1}^r \theta_i \tau_i^- = \sum_{i=1}^r \theta_i (\tau_i^+ + \tau_i^-) = 1$. The limit in (1.4.50) equals (1.4.53). Now, we use the fact that $\nabla_x \wp(\bar{x}, \bar{s} + \gamma_i^+ w_i)$ and $\nabla_x \wp(\bar{x}, \bar{s} + \gamma_i^- w_i)$ are contained in $\partial v(\bar{x})$ (see Lemma 1.4.12) implying that (1.4.53) is a convex combination of elements of $\partial v(\bar{x})$. Hence, (1.4.47), (1.4.48) and (1.4.50) yield $z \in \partial v(\bar{x})$.
(b) If $x^k \to x$ and ν_k, $\eta_k \to 0^+$ as $k \to \infty$, then the proof of the first assertion and (1.4.32), imply the existence of a converging subsequence of $(\nabla_x \widetilde{v}(x^k; \nu_k, \eta_k))$. □

Theorems 1.4.9 and 1.4.10 imply that $\widetilde{\psi}_j : \mathbb{R}^n \times \mathbb{R}^2_{++} \to \mathbb{R}$ defined by

$$\widetilde{\psi}_j(x;\nu,\eta) = -\min_{\widetilde{s}\in\mathbb{R}^{p+2}} \left\{ (1/2)\widetilde{s}^T \widetilde{H}_{\eta,j}(x)\widetilde{s} + \widetilde{g}_{\nu,j}(x)^T\widetilde{s} : \ \|\widetilde{s}\|_2 \le \Delta \right\}, \tag{1.4.54}$$

is a smoothing function for the function ψ_j defined in (1.1.6). Here $g_j(x) = -\bar{\Sigma}^{1/2} b_j(x)$, and $H_j(x) = -\bar{\Sigma}^{1/2} C_j(x) \bar{\Sigma}^{1/2}$. Moreover, $\widetilde{H}_{\eta,j}$ and $\widetilde{g}_{\nu,j}$ are defined as in (1.4.23) with H and g replaced by H_j and g_j, respectively. The representation of $\widetilde{\psi}_j$ results from (1.1.6) being transformed into the TRP (1.4.1) using $d \mapsto s = \bar{\Sigma}^{-1/2} d$.

Theorem 1.4.13 ([234, Thm. 5.13]). *Let $\bar{\Sigma} \in \mathbb{S}^p_{++}$, $q \ge 1$ and $j \in J$. Suppose that $a_j : \mathbb{R}^n \to \mathbb{R}$, $b_j : \mathbb{R}^n \to \mathbb{R}^p$ and $C_j : \mathbb{R}^n \to \mathbb{R}^p$ are q-times continuously differentiable. Then the following conditions hold true:*

(a) *The function $\widetilde{\psi}_j$ defined in (1.4.54) is a smoothing function for ψ_j, $\widetilde{\psi}_j(\cdot;\nu,\eta)$ is q-times continuously differentiable for each ν, $\eta > 0$, and gradient consistency holds.*

(b) *If $(x^k) \subset \mathbb{R}^n$ fulfills $x^k \to x \in \mathbb{R}^n$ as $k \to \infty$, and (ν_k), $(\eta_k) \subset \mathbb{R}_{++}$ converge 0 as $k \to \infty$, then there exists a convergent subsequence $(\nabla_x \widetilde{\psi}_j(x^k;\nu_k,\eta_k))_K$ of $(\nabla_x \widetilde{\psi}(x^k;\nu_k,\eta_k))$.*

The computational cost of evaluating (1.4.54) is essentially the same as the evaluation of (1.1.6) since the Hessian matrix $\widetilde{H}_{\eta,j}(x)$ defined in (1.4.23) is a block-diagonal matrix for $x \in \mathbb{R}^n$ implying that our smoothing approach is tractable both theoretically and practically.

1.5 Convergence of the Smoothing Method

We show that the accumulation points of the sequence of approximate KKT-tuples generated by Algorithm 1 are KKT-tuples of the approximated DROP (1.2.2) under mild assumptions. For each $j \in J$, we define the smoothing function $\widetilde{F}_j : \mathbb{R}^n \times \mathbb{R}^3_{++} \to \mathbb{R}$ for F_j by

$$\widetilde{F}_j(x;t) = a_j(x) + \widetilde{\varphi}_j(x;\tau) + \widetilde{\psi}_j(x;\nu,\eta). \tag{1.5.1}$$

Here $t = (\tau,\nu,\eta)$. We recall that $F_j : \mathbb{R}^n \to \mathbb{R}$ is given by $F_j(x) = a_j(x) + \varphi_j(x) + \psi_j(x)$ (see (1.2.1)), where $\widetilde{\varphi}_j$ and $\widetilde{\psi}_j$ are defined in (1.3.4) and (1.4.54), respectively. Under suitable assumptions on the approximated DROP (1.1.4), the smoothed DROP (1.2.3) has feasible points.

Proposition 1.5.1 ([234, Prop. 6.1]). *Let the hypotheses of Theorems 1.3.2 and 1.4.13 hold for each $j \in J \setminus \{0\}$. Suppose that $z \in \mathbb{R}^n$ is strictly feasible point for (1.2.2), that is, $z \in \mathbb{R}^n$ satisfies $F_j(z) < 0$ for each $j \in J \setminus \{0\}$. Then, for all sufficiently small $t > 0$, z is strictly feasible for (1.2.3).*

Proof. Theorems 1.3.2 and 1.4.13, and (1.2.1) and (1.5.1) yield, for each $j \in J \setminus \{0\}$,

$$\widetilde{F}_j(z;t) = a_j(x) + \widetilde{\varphi}_j(z;\tau) + \widetilde{\psi}_j(z;\nu,\eta) \to F_j(z) \quad \text{as} \quad t = (\tau,\nu,\eta) \to 0^+.$$

Hence, for all sufficiently small $t > 0$, the point z is strictly feasible for (1.2.3). □

Theorem 1.5.2 provides a global convergence result for Algorithm 1.

Theorem 1.5.2 ([234, Thm. 6.2]). *Let the conditions of Theorems 1.3.2 and 1.4.13 hold for each $j \in J$. Choose $\varepsilon_{\min}$, $t_{\min} = 0$. Suppose that the sequence $(x^k, \vartheta^k)_{\mathbb{N}_0}$ is generated by Algorithm 1. Then each accumulation point of $(x^k, \vartheta^k)_{\mathbb{N}_0}$ is a KKT-point of (1.2.2).*

Proof. Let $(\bar{x}, \bar{\vartheta})$ be an accumulation point of $(x^k, \vartheta^k)_{\mathbb{N}_0}$. Then there exists a subsequence $(x^k, \vartheta^k)_K$ of $(x^k, \vartheta^k)_{\mathbb{N}_0}$ converging to $(\bar{x}, \bar{\vartheta})$ as $K \ni k \to \infty$. Furthermore $0 \leq \chi(x^k, \vartheta^k; t^k) \leq \varepsilon_k$ for all $k \in \mathbb{N}_0$, where the criticality measure χ is defined in (1.2.4). Since $\varepsilon_k \to 0$ as $k \to \infty$, (1.5.1), Theorems 1.3.2 and 1.4.13 give, for each $j \in J \setminus \{0\}$,

$$\varepsilon_k \geq |\min\{-\widetilde{F}_j(x^k; t^k), \vartheta_j^k\}| \to |\min\{-F_j(\bar{x}), \bar{\vartheta}_j\}| = 0 \ \text{ as } \ K \ni k \to \infty.$$

Since $(a, b) \mapsto \min\{a, b\}$ is a complementarity function (see, e.g., [321, sect. 1.3]), we have $\bar{\vartheta}_j F_j(\bar{x}) = 0$, $F_j(\bar{x}) \leq 0$, and $\bar{\vartheta}_j \geq 0$ for all $j \in J \setminus \{0\}$. Owing to Theorems 1.3.2 and 1.4.13, we can assume without loss of generality that the sequences $(\nabla_x \widetilde{\varphi}_j(x^k; \tau_k))_K$ and $(\nabla_x \widetilde{\psi}_j(x^k; \nu_k, \eta_k))_K$ for $j \in J$ are convergent. Hence, for each $j \in J$, there exist v_j, $w_j \in \mathbb{R}^n$ such that

$$\nabla_x \widetilde{\varphi}_j(x^k; \tau_k) \to v_j \quad \text{and} \quad \nabla_x \widetilde{\psi}_j(x^k; \nu_k, \eta_k) \to w_j \quad \text{as} \quad K \ni k \to \infty.$$

Now, fix $j \in J$. We verify that $\nabla a_j(\bar{x}) + v_j + w_j \in \partial F_j(\bar{x})$. Theorems 1.3.2 and 1.4.13 give $v_j \in \partial \varphi_j(\bar{x})$ and $w_j \in \partial \psi_j(\bar{x})$ because of gradient consistency. Since φ_j and ψ_j are optimal value functions, [78, Thm. 2.1] implies their regularity. Moreover, sums of regular functions are regular [79, Prop. 2.3.6]. Combined with the continuous differentiability of a_j, and the sum rule [79, Cor. 3 on p. 40], we find that $\partial F_j(\bar{x}) = \nabla a_j(\bar{x}) + \partial \varphi_j(\bar{x}) + \partial \psi_j(\bar{x})$. Hence $\nabla a_j(\bar{x}) + v_j + w_j \in \partial F_j(\bar{x})$ and, consequently,

$$\nabla a_0(\bar{x}) + v_0 + w_0 + \sum_{j \in J \setminus \{0\}} \bar{\vartheta}_j (\nabla a_j(\bar{x}) + v_j + w_j) \in \partial F_0(\bar{x}) + \sum_{j \in J \setminus \{0\}} \bar{\vartheta}_j \partial F_j(\bar{x}).$$

Moreover, $\chi(x^k, \vartheta^k; t^k) \to 0$ as $k \to \infty$ ensures

$$\nabla_x \widetilde{F}_0(x^k; t^k) + \sum_{j \in J \setminus \{0\}} (\vartheta^k)_j \nabla_x \widetilde{F}_j(x^k; t^k) \to 0 \quad \text{as} \quad K \ni k \to \infty.$$

Putting together the statements, we conclude that $0 \in \partial F_0(\bar{x}) + \sum_{j \in J \setminus \{0\}} \bar{\vartheta}_j \partial F_j(\bar{x})$. □

If we only assume that (x^k) has a convergent subsequence, we may need to impose a suitable constraint qualification for the approximated DROP (1.2.2) in order to deduce the convergence of a subsequence of (ϑ^k) [349, Thm. 3.2]. Moreover, the existence of KKT-tuples for the smoothed DROP (1.2.3) may be shown under suitable constraint qualifications for the approximated DROP (1.2.2); see [349, sects. 2 and 3].

1.6 Numerical Simulations

We construct DROPs from the Moré–Garbow–Hillstrom test set [237] consisting of unconstrained NLPs. In order to obtain DROPs, we model design variables as uncertain. This type of uncertainty is referred to as implementation errors in the literature on robust optimization [23, p. 4], [31, p. 166]. Robust nonlinear optimization with implementation errors is considered in, for example, [31, 32, 209, 162].

We consider the nonlinear DROP

$$\min_{x \in \mathbb{R}^n} \sup_{P \in \mathcal{P}_\epsilon} \mathbb{E}_P[f_0(x + \xi)], \tag{1.6.1}$$

where $\mathcal{P}_\epsilon$ is defined in (1.1.2) with $\bar{\mu} = 0$, $\Delta = \sqrt{\epsilon}$, $\bar{\Sigma}_0 = 0$, and $\bar{\Sigma} = \bar{\Sigma}_1 = \epsilon I$. We choose $\epsilon \in \{10^{-3}, 10^{-2}\}$ and refer the reader to section 1.8.1 for a description of how we selected problems from the Moré–Garbow–Hillstrom test set. We consider the approximated DROP

$$\min_{x \in \mathbb{R}^n} F_0(x), \tag{1.6.2}$$

where F_0 is defined in (1.2.1), and $a_0(x) = f_0(x)$, $b_0(x) = \nabla f_0(x)$, and $C_0(x) = \nabla^2 f_0(x)$ are chosen in (1.1.6) and (1.1.5).
One goal of our numerical tests is to show that stationary points of (1.6.2) are more resilient to distributional uncertainty than those of the nominal problem

$$\min_{x \in \mathbb{R}^n} \; f_0(x), \tag{1.6.3}$$

and of the sample average approximation (SAA) of the stochastic program

$$\min_{x \in \mathbb{R}^n} \; \mathbb{E}_{\bar{P}_\epsilon}[f_0(x + \xi)], \tag{1.6.4}$$

with $\bar{P}_\epsilon = \mathcal{N}(0, (\epsilon/10)I)$, despite the fact that we approximate the DROP (1.6.1) by the DROP (1.6.2). We compare these stationary points in section 1.6.4. We chose the nominal distribution $\bar{P}_\epsilon = \mathcal{N}(0, (\epsilon/10)I)$ to mimic the setup considered by Delage and Ye [94, sect. 4.3].
A further goal of our experiments is to show that Algorithm 1 is an efficient method to compute stationary points of (1.6.2). We compare the performance of Algorithm 1 with the bundle method `MPBNGC` [223, 224] applied to (1.6.2), and the NSDP solver `PENLAB` [110] applied to

$$\begin{array}{ll} \min\limits_{x \in \mathbb{R}^n, \gamma \in \mathbb{R}, \lambda \in \mathbb{R}_+, y \in \mathbb{R}^p, \Lambda, \Upsilon \in \mathbb{S}^p_+} & a_0(x) - (1/2)\gamma + (1/2) I \bullet \Upsilon \\ \text{s.t.} \quad \begin{bmatrix} \lambda I + \Upsilon - \Lambda & y \\ y^T & -\lambda \Delta^2 - \gamma \end{bmatrix} \succcurlyeq 0, & \begin{array}{c} \operatorname{svec}(\Upsilon - \Lambda + \bar{\Sigma}^{1/2} C_0(x) \bar{\Sigma}^{1/2}) = 0, \\ y + \bar{\Sigma}^{1/2} b_0(x) = 0, \end{array} \end{array} \tag{1.6.5}$$

where $\operatorname{svec} : \mathbb{S}^p \to \mathbb{R}^{p(p+1)/2}$ transforms the lower triangular part of a symmetric matrix into a vector. Using Proposition 1.8.3, we find that the NSDP (1.6.5) is equivalent to the approximated DROP (1.6.2). Different from (1.8.12), we have introduced a slack variable in (1.6.5) and "preconditioned" b_0 and C_0 using $\bar{\Sigma}^{1/2}$. `PENLAB` [110] preforms better when applied to the NSDP reformulation (1.6.5) than when applied to (1.8.12).
Out of eight solvers for nonsmooth, nonconvex optimization, the decision tree for nonsmooth optimization software, `Solver-o-matic` [174], recommended the use of `MPBNGC` for the approximated DROP (1.6.2).

1.6.1 Implementation Details

We provide implementation details for Algorithm 1, and for the applications of `MPBNGC` to (1.6.2) and `PENLAB` to (1.6.5). Algorithm 1 was implemented in `Julia` [35] using `Ipopt` [336] and its `Julia` interface `Ipopt.jl`. We chose the same stopping criterion for each iteration of Algorithm 1. We used the default settings of `Ipopt`, with the exception of modifying the overall termination tolerance `tol`. The gradient of the smoothing functions $\widetilde{\varphi}_0$ (see (1.3.4)) and $\widetilde{\psi}_0$ (see (1.4.54)) were computed with the formulas (1.3.7) and (1.4.32), respectively, and `Ipopt` was used with L-BFGS. We chose $\nu_{\min} = 10^{-8}$, $\eta_{\min}$, $\tau_{\min} = \sqrt{\nu_{\min}}$, η_0, $\tau_0 = \sqrt{\nu_0}$, $\nu_{k+1} = \min\{\rho^2 \nu_k, \nu_{\min}\}$, and η_{k+1}, $\tau_{k+1} = \min\{\rho \eta_k, \nu_{\min}\}$, where $\nu_0 > 0$, $\rho = 0.1$. For $\texttt{tol} = 10^{-4}$ and $\nu_0 = 0.1$, the above choices of the smoothing parameters are motivated by Theorems 1.3.2 and 1.4.6. Evaluating the smoothing function $\widetilde{F}_0$ (see (1.5.1)) of F_0 (see (1.6.2)) at (x, t) requires $f_0(x)$ (see (1.6.3)), $\nabla f_0(x)$ and $\nabla^2 f_0(x)$. To obtain $\nabla_x \widetilde{F}_0(x; t)$, we computed the gradients of $x \mapsto \widetilde{s}^T \nabla^2 f_0(x) \widetilde{s}$, where $\widetilde{s}$ are the first p components of the optimal solution of the TRP (1.4.54), and of two mapping of the form $x \mapsto \nabla^2 f_0(x) \bullet R$, where $R \in \mathbb{S}^p$; see (1.3.7) and (1.4.32). To initialize the solution of the smoothed problem (1.2.3) in the $(k+1)$th iteration of Algorithm 1, we used the approximate stationary point obtained in the kth iteration.
We implemented and used a `Julia` interface for `MPBNGC`. For `MPBNGC`, we used the same setup as in [223, sect. 6], except that we chose different termination tolerances and set `GAM` $= 0.5$

for each test problems. Since MPBNGC is a bundle method, it requires function and subgradient evaluations of $F_0 = a_0 + \varphi_0 + \psi_0$. We exploited the regularity of the optimal value functions φ_0 and ψ_0 (see (1.1.5) and (1.1.6)) and computed subgradients of F_0 via the sum rule [79, Cor. on p. 39]. We obtained subgradients of φ_0 and of ψ_0 using (1.3.10) and (1.4.47), respectively.
PENLAB requires first and second derivatives of the objective function and each constraint function of the NSDP (1.6.5). For PENLAB, we computed derivatives of f_0 up to fourth order using the automatic differentiation tool ADiGator [341]. We excluded the test function mgh35 from the tests with PENLAB as ADiGator does not support automatic differentiation for the objective function of this problem. We chose the same initial values for x that we passed to Algorithm 1. We obtained the remaining initial points for the variables in (1.6.5) by applying PENLAB to (1.6.5) for fixed x. We chose outer_stop_limit=kkt_stop_limit. For the remaining settings, we used PENLAB's default.
We scaled f_0 using Ipopt's the gradient scaling, which is described in [336, sect. 3.8]. We chose x_N^* as initial value for Algorithm 1 and MPBNGC, where x_N^* is the stationary point computed by Ipopt for the nominal problem (1.6.3) with termination tolerance 10^{-5} and the above settings. The TRPs (1.1.6) and (1.4.54) were solved using [238, Alg. 3.14]. For the Julia codes, we computed derivatives of f_0 using the automatic differentiation package ForwardDiff [266]. We took advantage of the fact that the DROP (1.6.1) models uncertain decision variables, and made use of $\nabla f_0 = \nabla a_0 = b_0$ and $\nabla b_0 = C_0$.

1.6.2 Comparison of the Smoothing Method with MPBNGC and PENLAB

We compare the performance of the smoothing method, Algorithm 1, with that of MPBNGC and of PENLAB in terms of evaluations of f_0 and its derivatives. The termination criteria of Ipopt (see [336, sect. 2.1]), MPBNGC (see [224, sect. 3.3]), and PENLAB (see [110, Alg. 1]) are different. To be able to make a fair comparison or nearly so, we applied Ipopt to each nominal problem (1.6.3) with known exact solution (see [237, sect. 3]) and computed the median of the absolute errors of the final objective function values returned by Ipopt and the true ones with tol $= 10^{-2}, 10^{-4}$. Then, we applied MPBNGC and PENLAB to the same problems with termination tolerances $10^{-1}, 10^{-2}, \ldots, 10^{-10}$, and from this list computed the largest ones such that we obtained the same order of magnitude of the errors as with Ipopt for the tolerances $10^{-2}, 10^{-4}$. The corresponding criteria are $10^{-4}, 10^{-8}$ for MPBNGC, and $10^{-2}, 10^{-5}$ for PENLAB. This type of "calibration" tries to ensure that stationary points obtained via Algorithm 1, MPBNGC, and PENLAB are of similar accuracy.
Ipopt preforms many line search calls in Algorithm 1. To account for inexact evaluations of $\varphi = \widetilde{F}_0(\cdot; t^k)$ resulting from the inexact solution of the TRPs (1.4.54), we provide, in addition to using Ipopt's line search, numerical results using the modified line search

$$\varphi(x^i + \alpha_i d^i) - 10\varepsilon_{\text{mach}}|\varphi(x^i)| - 2\sqrt{\varepsilon_{\text{mach}}}|\varphi(x^i)| \leq \varphi(x^i) + \kappa\alpha_i \nabla\varphi(x^i)^T d^i,$$

where $\varepsilon_{\text{mach}} \approx 10^{-16}$ is the machine precision, $t^k = (\nu_k, \eta_k, \tau_k)$, $\kappa \in (0, 1/2)$ is a default value of a parameter of Ipopt, x^i is the current iterate, d^i is a step, and $\alpha_i > 0$ is a line search parameter [336, sect. 3.10]. The modified line search differs from that used by Ipopt [336, sect. 3.10] in that we subtract $2\sqrt{\varepsilon_{\text{mach}}}|\varphi(x^i)|$.[2] Ipopt's line search is only designed to account for round-off errors via the term $-10\varepsilon_{\text{mach}}|\varphi(x^i)|$ [336, sect. 3.10]. For each Moré–Garbow–Hillstrom test problem, Ipopt using the modified line search has similar convergence behavior as Ipopt using its line search.

[2] We assume that we can evaluate $\varphi = \widetilde{F}_0(\cdot; t^k)$ with approximately eight significant figures. Define $u = \varphi(x^i + \alpha_i d^i)$ and $w = \varphi(x^i)$ and denote by $\widehat{u}$ and $\widehat{w}$ their numerical approximation, respectively. We have $|u - \widehat{u}| \approx \sqrt{\varepsilon_{\text{mach}}}|u|$ and $|w - \widehat{w}| \approx \sqrt{\varepsilon_{\text{mach}}}|w|$. If $|u| \approx |w|$, then $|u - \widehat{u} - (w - \widehat{w})| \approx 2\sqrt{\varepsilon_{\text{mach}}}|u|$ motivating the term $2\sqrt{\varepsilon_{\text{mach}}}|\varphi(x^i)|$ in the modified line search.

TABLE 1.1: *Median number of evaluations of derivatives of f_0, $x \mapsto d^T\nabla^2 f_0(x)d$, $d \in \mathbb{R}^p$, and of $x \mapsto \nabla^2 f_0(x) \bullet R$, $R \in \mathbb{S}^p$, used by Algorithm 1, `MPBNGC`, and `PENLAB` with $\epsilon = 10^{-3}$. Each number is rounded to its nearest integer. `SgM(tol,` ν_0`)` refers to Algorithm 1 with termination tolerance `tol` and initial smoothing parameter ν_0 using `Ipopt` as NLP solver. `mSgM(tol,` ν_0`)` refers to `SgM(tol,` ν_0`)` using `Ipopt` with the modified line search.*

Method	$\#f_0$, $\#\nabla f_0$, $\#\nabla^2 f_0$	$\#\nabla(d^T\nabla^2 f_0 d)$	$\#\nabla(\nabla^2 f_0 \bullet R)$	$\#\mathrm{D}^3 f_0$	$\#\mathrm{D}^4 f_0$
SgM(10^{-2}, 10^{-1})	76	15	30		
SgM(10^{-2}, 10^{-3})	61	15	30		
mSgM(10^{-2}, 10^{-3})	52	12	25		
MPBNGC	24	24	24		
PENLAB	54			29	23
SgM(10^{-4}, 10^{-1})	120	37	75		
SgM(10^{-4}, 10^{-3})	101	32	64		
mSgM(10^{-4}, 10^{-3})	85	27	54		
MPBNGC	69	69	69		
PENLAB	88			59	46

We report the median number of evaluations of f_0, of its derivatives, and of the derivatives $\nabla(d^T\nabla^2 f_0(\cdot)d)$, $d \in \mathbb{R}^p$, and $\nabla(\nabla^2 f_0(\cdot) \bullet R)$, $R \in \mathbb{S}^p$, used by Algorithm 1, `MPBNGC`, and `PENLAB` with $\epsilon = 10^{-3}$ in Table 1.1. For each selected test problem, the number of evaluations used in Algorithm 1 is the sum of all evaluations of the inner iterations. We chose initial smoothing parameters $\nu_0 = 10^{-1}$, 10^{-3}, and $\eta_0, \tau_0 = \sqrt{\nu_0}$. Moreover, we used the termination tolerances `tol` $= 10^{-2}$, 10^{-4} for Algorithm 1, and for `MPBNGC` and `PENLAB` the corresponding ones as stated above. Instead of evaluating gradient of $x \mapsto d^T\nabla^2 f_0(x)d$, and of $x \mapsto \nabla^2 f_0(x) \bullet R$ for two $R \in \mathbb{S}^p$, we could have computed $\mathrm{D}^3 f_0$ once.
Table 1.1 indicates that Algorithm 1 requires about half as many gradient evaluations of $\widetilde{F}_0$ as `MPBNGC` requires subgradient evaluations of F_0. `PENLAB` requires, as opposed to Algorithm 1 and `MPBNGC`, third and fourth derivatives of f_0. Combined with the results displayed in Table 1.1, we may conclude that `PENLAB` is the most expensive method in terms of evaluations of f_0 and of its derivatives. Table 1.1 indicates that small initial smoothing parameters can be beneficial, as they result in fewer evaluations. It also indicates that using `Ipopt` with the modified line search can lead to fewer function evaluations of $\widetilde{F}_0(\cdot; t^k)$.

1.6.3 Details on Performance of Smoothing Method

We discuss the performance of Algorithm 1 as a smoothing method with `tol` $= 10^{-4}$ and $\nu_0 = 0.1$. Table 1.2 displays the number of inner and outer iterations for mgh01 and mgh03. Moreover, it displays the KKT-error, the distance of the stationary point of the current iteration to that of the previous iteration and the smoothing parameter ν_k for each outer iteration k of Algorithm 1. Empirically the distance of subsequent stationary points (1.2.3) converges to zero and the number of inner iterations decreases monotonically, indicating that the smoothing method is efficient.
The solution of the TRPs (1.4.54) using the Moré–Sorensen algorithm, [238, Alg. 3.14], for all iterations of Algorithm 1 required fewer than six iterations which is in accordance with the results presented in [238, sect. 5].

1.6.4 Comparison of Stationary Points

For each selected problem, we compare the stationary points x^*_{DR} of (1.6.2) computed with Algorithm 1 using `tol` $= 10^{-4}$ and $\nu_0 = 0.1$, x^*_N of (1.6.3) and x^*_S of an SAA of (1.6.4) using

Table 1.2: *For each outer iteration of Algorithm 1 applied to* (1.6.1) *and* $\epsilon = 10^{-3}$*, the number of iterations required to compute a stationary point of* (1.2.3)*, the final KKT-error, the relative distance of the initial point and the stationary point, and the value of the smoothing parameter* ν_k *are displayed.*

Problem	k	#-iter	KKT-error	$\frac{\|x^k - x^{k-1}\|_2}{\max\{1, \|x^{k-1}\|_2\}}$	ν_k
mgh01	0	17	$2.272 \cdot 10^{-7}$	0.3329	0.1
	1	10	$2.756 \cdot 10^{-6}$	$9.633 \cdot 10^{-2}$	10^{-3}
	2	2	$4.34 \cdot 10^{-5}$	$5.013 \cdot 10^{-5}$	10^{-5}
	3	2	$6.355 \cdot 10^{-5}$	$3.975 \cdot 10^{-5}$	10^{-7}
	4	2	$9.689 \cdot 10^{-5}$	$3.345 \cdot 10^{-5}$	10^{-8}
mgh03	0	25	$7.654 \cdot 10^{-5}$	0.9994	0.1
	1	7	$6.938 \cdot 10^{-7}$	$5.542 \cdot 10^{-5}$	10^{-3}
	2	7	$2.303 \cdot 10^{-7}$	$5.495 \cdot 10^{-6}$	10^{-5}
	3	5	$4.997 \cdot 10^{-7}$	$5.494 \cdot 10^{-7}$	10^{-7}
	4	5	$1.015 \cdot 10^{-6}$	$4.07 \cdot 10^{-8}$	10^{-8}

the following two quantities:

$$V_{\mathbb{E}}(x) = \max_{1 \le i \le 10} \mathbb{E}_{P_i}[f_0(x + \xi_i)] \quad \text{and} \quad V_{\mathrm{StD}}(x) = \max_{1 \le i \le 10} \mathrm{StD}_{P_i}[f_0(x + \xi_i)], \tag{1.6.6}$$

where $P_i = \mathcal{N}(\mu_i, \sigma_i^2 I) \in \mathcal{P}_\epsilon$, and μ_i and σ_i are independent and uniformly distributed on $\{\, \mu \in \mathbb{R}^p : \|\mu\|_2 \le \Delta \,\}$ and $\{\, \sigma \in \mathbb{R} : 0 \le \sigma^2 \le \epsilon \,\}$, respectively. Here, StD is the standard deviation. We approximated expected values using empirical averages with 1000 independent samples.

The quantities in (1.6.6) mimic the maximum mean and standard deviations of repeated implementations of x, and $V_{\mathbb{E}}$ is a lower bound on the objective function of (1.6.1). We computed the stationary points x_N^* and x_S^* using `Ipopt` with `tol` $= 10^{-5}$ and exact Hessian information for nominal and stochastic programs. Tables 1.3 and 1.4 display $V_{\mathbb{E}}(x)$ and $V_{\mathrm{StD}}(x)$ for $x \in \{x_{\mathrm{DR}}^*, x_N^*, x_S^*\}$, and $\epsilon \in \{10^{-3}, 10^{-2}\}$. In most cases, the distributionally robust stationary points have lower mean and standard deviation than the nominal and the stochastic stationary points. We conclude that the stationary points of the approximated DROP are more resilient to distributional uncertainty than those of the nominal and stochastic problems, for many test problems.

The problems mgh33 and mgh34 are quadratic w.r.t. ξ (see [237, sect. 3]). Consequently, the approximation scheme is exact, that is, the DROP (1.6.1) is equivalent to the approximated DROP (1.6.2). For the problems mgh10, mgh11 and mgh17, we obtained very different orders of magnitude of $V_{\mathbb{E}}(x)$, and of $V_{\mathrm{StD}}(x)$ for $x \in \{x_N^*, x_{\mathrm{DR}}^*, x_S^*\}$, resulting from exponential terms in the corresponding objective functions [237, sect. 3].

Table 1.5 lists the median number of corresponding objective function, gradient and Hessian evaluations used by `Ipopt` to compute a stationary point of (1.6.2) using Algorithm 1, of (1.6.3), and of the sample average approximation of (1.6.4).

TABLE 1.3: *Quantities $V_{\mathbb{E}}$ and V_{StD} (see (1.6.6)) evaluated at x_N^*, x_{DR}^*, x_S^* for $\epsilon = 10^{-3}$.*

Problem	$V_{\mathbb{E}}(x_N^*)$	$V_{\mathbb{E}}(x_{\mathrm{DR}}^*)$	$V_{\mathbb{E}}(x_S^*)$	$V_{\mathrm{StD}}(x_N^*)$	$V_{\mathrm{StD}}(x_{\mathrm{DR}}^*)$	$V_{\mathrm{StD}}(x_S^*)$
mgh01	0.1867	0.1536	0.1761	0.2559	0.1528	0.2399
mgh03	$3.175 \cdot 10^{6}$	$3.135 \cdot 10^{1}$	$2.899 \cdot 10^{1}$	$4.507 \cdot 10^{6}$	$8.083 \cdot 10^{1}$	$7.328 \cdot 10^{1}$
mgh04	$3.756 \cdot 10^{8}$	$3.754 \cdot 10^{8}$	$3.754 \cdot 10^{8}$	$5.256 \cdot 10^{8}$	$5.246 \cdot 10^{8}$	$5.252 \cdot 10^{8}$
mgh06	$1.884 \cdot 10^{2}$	$1.798 \cdot 10^{2}$	$1.863 \cdot 10^{2}$	$1.076 \cdot 10^{2}$	$8.388 \cdot 10^{1}$	$1.028 \cdot 10^{2}$
mgh07	0.1778	0.1778	0.1779	0.2089	0.2086	0.209
mgh10	$9.626 \cdot 10^{10}$	$1.356 \cdot 10^{6}$	$2.134 \cdot 10^{6}$	$1.303 \cdot 10^{11}$	$3.482 \cdot 10^{5}$	$1.993 \cdot 10^{6}$
mgh11	$6.237 \cdot 10^{278}$	$3.283 \cdot 10^{1}$	$2.258 \cdot 10^{133}$	∞	0.8311	$7.134 \cdot 10^{134}$
mgh13	$4.387 \cdot 10^{-2}$	$4.387 \cdot 10^{-2}$	$4.385 \cdot 10^{-2}$	$5.662 \cdot 10^{-2}$	$5.662 \cdot 10^{-2}$	$5.66 \cdot 10^{-2}$
mgh14	0.7525	0.7492	0.752	0.7223	0.7144	0.7229
mgh17	$7.9421 \cdot 10^{17}$	1.133	$1.735 \cdot 10^{11}$	$2.19 \cdot 10^{19}$	$3.551 \cdot 10^{-2}$	$4.959 \cdot 10^{12}$
mgh20	0.1318	0.1291	0.1309	0.1461	0.1425	0.1453
mgh21	3.92	3.19	3.702	1.723	1.045	1.621
mgh22	0.2164	0.2163	0.2163	0.1219	0.1219	0.1219
mgh25	0.3078	0.3078	0.3073	0.6784	0.6784	0.6768
mgh27	$4.855 \cdot 10^{-2}$	$4.853 \cdot 10^{-2}$	$4.85 \cdot 10^{-2}$	$6.864 \cdot 10^{-2}$	$6.854 \cdot 10^{-2}$	$6.859 \cdot 10^{-2}$
mgh30	0.1408	0.1406	0.1408	$7.52 \cdot 10^{-2}$	$7.485 \cdot 10^{-2}$	$7.519 \cdot 10^{-2}$
mgh31	0.194	0.1924	0.1936	$9.437 \cdot 10^{-2}$	$8.959 \cdot 10^{-2}$	$9.399 \cdot 10^{-2}$
mgh33	$4.514 \cdot 10^{2}$	$4.514 \cdot 10^{2}$	$4.508 \cdot 10^{2}$	$6.369 \cdot 10^{2}$	$6.369 \cdot 10^{2}$	$6.365 \cdot 10^{2}$
mgh34	$2.394 \cdot 10^{2}$	$2.394 \cdot 10^{2}$	$2.391 \cdot 10^{2}$	$3.203 \cdot 10^{2}$	$3.203 \cdot 10^{2}$	$3.204 \cdot 10^{2}$
mgh35	$6.772 \cdot 10^{-2}$	$5.266 \cdot 10^{-2}$	0.1244	0.3531	$2.726 \cdot 10^{-2}$	1.383

TABLE 1.4: *Quantities $V_{\mathbb{E}}$ and V_{StD} (see (1.6.6)) evaluated at x_N^*, x_{DR}^*, x_S^* for $\epsilon = 10^{-2}$.*

Problem	$V_{\mathbb{E}}(x_N^*)$	$V_{\mathbb{E}}(x_{\mathrm{DR}}^*)$	$V_{\mathbb{E}}(x_S^*)$	$V_{\mathrm{StD}}(x_N^*)$	$V_{\mathrm{StD}}(x_{\mathrm{DR}}^*)$	$V_{\mathrm{StD}}(x_S^*)$
mgh01	1.866	0.7566	1.174	2.581	0.5774	1.548
mgh03	$3.178 \cdot 10^{7}$	$2.829 \cdot 10^{3}$	$2.81 \cdot 10^{3}$	$4.511 \cdot 10^{7}$	$7.443 \cdot 10^{3}$	$7.414 \cdot 10^{3}$
mgh04	$3.752 \cdot 10^{9}$	$3.728 \cdot 10^{9}$	$3.744 \cdot 10^{9}$	$5.246 \cdot 10^{9}$	$5.142 \cdot 10^{9}$	$5.233 \cdot 10^{9}$
mgh06	$1.903 \cdot 10^{3}$	$8.186 \cdot 10^{2}$	$1.528 \cdot 10^{3}$	$5.762 \cdot 10^{3}$	$2.055 \cdot 10^{3}$	$4.626 \cdot 10^{3}$
mgh07	1.793	1.781	1.792	2.129	2.09	2.125
mgh10	$9.616 \cdot 10^{11}$	$9.616 \cdot 10^{11}$	$3.502 \cdot 10^{6}$	$1.294 \cdot 10^{12}$	$1.294 \cdot 10^{12}$	$3.261 \cdot 10^{6}$
mgh11	$8.08 \cdot 10^{256}$	$3.283 \cdot 10^{1}$	$7.044 \cdot 10^{129}$	∞	0.7971	$2.228 \cdot 10^{131}$
mgh13	0.4413	0.4413	0.4412	0.5675	0.5675	0.5674
mgh14	7.537	7.262	7.466	7.318	6.657	7.245
mgh17	$3.857 \cdot 10^{68}$	1.374	$8.32 \cdot 10^{46}$	$1.135 \cdot 10^{70}$	0.3561	$2.527 \cdot 10^{48}$
mgh20	1.321	1.262	1.279	1.487	1.427	1.445
mgh21	$3.941 \cdot 10^{1}$	$1.556 \cdot 10^{1}$	$2.496 \cdot 10^{1}$	$1.758 \cdot 10^{1}$	3.972	$1.083 \cdot 10^{1}$
mgh22	2.184	2.183	2.183	1.219	1.219	1.219
mgh25	$1.647 \cdot 10^{1}$	$1.647 \cdot 10^{1}$	$1.643 \cdot 10^{1}$	$5.02 \cdot 10^{1}$	$5.021 \cdot 10^{1}$	$5.0 \cdot 10^{1}$
mgh27	0.4891	0.4876	0.488	0.7019	0.6896	0.6995
mgh30	1.408	1.382	1.402	0.7588	0.7236	0.755
mgh31	2.063	1.846	2.01	1.217	0.8228	1.166
mgh33	$4.516 \cdot 10^{3}$	$4.516 \cdot 10^{3}$	$4.506 \cdot 10^{3}$	$6.392 \cdot 10^{3}$	$6.392 \cdot 10^{3}$	$6.381 \cdot 10^{3}$
mgh34	$2.378 \cdot 10^{3}$	$2.378 \cdot 10^{3}$	$2.372 \cdot 10^{3}$	$3.207 \cdot 10^{3}$	$3.207 \cdot 10^{3}$	$3.204 \cdot 10^{3}$
mgh35	$1.221 \cdot 10^{3}$	$3.846 \cdot 10^{3}$	$3.717 \cdot 10^{2}$	$2.563 \cdot 10^{4}$	$4.247 \cdot 10^{4}$	$7.365 \cdot 10^{3}$

TABLE 1.5: *Median number of objective function, gradient, and Hessian evaluations required by* `Ipopt` *for the nominal (N), distributionally robust (DR) and stochastic optimization problem (S) of all selected test problems. The number of evaluations for the approximate DROPs (1.6.2) are the sum of all evaluations used within Algorithm 1.*

	N	DR	S	N	DR	S	N	S
ϵ	#-f_0	#-$\widetilde{F}_0$	#-f_0	#-∇f_0	#-$\nabla_x \widetilde{F}_0$	#-∇f_0	#-$\nabla^2 f_0$	#-$\nabla^2 f_0$
10^{-3}	14	120	$1.25 \cdot 10^{4}$	14	37.5	$1.2 \cdot 10^{4}$	13	$1.1 \cdot 10^{4}$
10^{-2}	14	190.5	$1.0 \cdot 10^{4}$	14	56	$1.0 \cdot 10^{4}$	13	$0.9 \cdot 10^{3}$

1.7 Conclusion and Discussion

We developed an approximation scheme for moment-based distributionally robust nonlinear optimization. Using second-order expansions of the parameterized objective function and each of the constraint functions, we obtained an approximated DROP defined by nonsmooth optimal value functions. We constructed smoothing functions for these nonsmooth functions which satisfy gradient consistency in section 1.2. Our approach allowed us to apply derivative-based optimization methods within our smoothing method, Algorithm 1. The global convergence of the smoothing method to stationary points of the approximated DROP was shown using the gradient consistency of the smoothing functions in section 1.5.
In section 1.6, we compared our algorithmic approach with the application of a bundle method to the approximated DROP and of a solver for NSDPs applied to an equivalent reformulation of it as a NSDP. Our numerical results indicate that the smoothing method, Algorithm 1, computes fewer derivatives than the bundle method `MPBNGC` evaluates subgradients. `PENLAB` applied to the NSDP reformulation (1.6.5) of the approximated DROP (1.6.2) requires the computation of derivatives of matrix-valued constraints, making it the most expensive method to compute stationary points of the approximated DROP (1.6.2).
Our scheme has the following features: (i) the number of constraints and of optimization variables of the approximated and smoothed DROPs is the same as for the nominal problem; (ii) mathematical programs with complementarity constraints and NSDPs are avoided; (iii) the smoothing functions can efficiently be evaluated using existing computer codes and satisfy gradient consistency; and (iv) within our smoothing method, Algorithm 1, any NLP solver can be applied to compute a sequence of approximate stationary points of the smoothed DROPs.
Our approximation scheme and algorithmic approach can also be used for a different ambiguity set. We define the ambiguity set $\widehat{\mathcal{P}}$ by

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

where $\Delta > 0$, $\sigma_0 < \sigma_1$, and $\widehat{\Sigma} \in \mathbb{S}^p_{++}$. As opposed to imposing an explicit bound on the covariance as in (1.1.2), the ambiguity set $\widehat{\mathcal{P}}$ restricts the second moments. Let $P \in \widehat{\mathcal{P}}$ and $x \in \mathbb{R}^n$ be arbitrary. Using (1.1.3), we have

$$\mathbb{E}_P[m_j(x,\xi)] = a_j(x) + b_j(x)^T(\mathbb{E}_P[\xi] - \bar{\mu}) + (1/2)C_j(x) \bullet \mathbb{E}_P[\xi\xi^T].$$

Combined with $\max_{d\in\mathbb{R}^p}\{\, b_j(x)^T(d-\bar{\mu}) :\ \|\widehat{\Sigma}^{-1/2}(d-\bar{\mu})\|_2 \leq \Delta\,\} = \|\widehat{\Sigma}^{1/2}b_j(x)\|_2$ [250, p. 90], we obtain, for each $x \in \mathbb{R}^n$,

$$\sup_{P\in\mathcal{P}} \mathbb{E}_P[m_j(x,\xi)] = a_j(x) + \|\widehat{\Sigma}^{1/2}b_j(x)\|_2 + \max_{\sigma_0\widehat{\Sigma}\preccurlyeq\Sigma\preccurlyeq\sigma_1\widehat{\Sigma}} \big\{\, (1/2)C_j(x) \bullet \Sigma \,\big\}.$$

The optimal value function is an instance of those considered in section 1.3.
Our scheme provides an alternative to those used, for example, in [181, 209, 199, 218, 158] for robust nonlinear optimization. An open research task is to compare our approximation scheme and algorithmic approach to those used in [181, 209, 199, 218, 158] for robust nonlinear optimization.
We used second-order Taylor's expansion about the nominal parameter of the parameterized objective and each of the constraint functions to derive accurate approximations. However, for some problems, more accurate surrogate functions may be desirable. For robust nonlinear optimization, Lass and Ulbrich [209] propose a strategy to adaptively shift the expansion point (see also [181, 180]). This approach may be extended to and used for nonlinear DROPs. If samples

TABLE 1.6: *The quantity $Z_\epsilon(x_N^*)$ defined in (1.8.1), the problems from the Moré–Garbow–Hillstrom test set with $Z_\epsilon(x_N^*)$ exceeding 1/10 for $\epsilon = 10^{-3}$, and the corresponding number of parameters p.*

Problem	p	$Z_\epsilon(x_N^*)$	Problem	p	$Z_\epsilon(x_N^*)$	Problem	p	$Z_\epsilon(x_N^*)$
mgh01	2	0.5778	mgh13	4	0.1262	mgh27	10	0.1403
mgh03	2	$1.006 \cdot 10^7$	mgh14	4	1.911	mgh30	10	0.2737
mgh04	2	$1.234 \cdot 10^9$	mgh17	5	$8.83 \cdot 10^{24}$	mgh31	10	0.3551
mgh06	2	1.903	mgh20	6	0.3353	mgh33	10	$5.505 \cdot 10^2$
mgh07	3	0.4727	mgh21	20	7.056	mgh34	10	$2.243 \cdot 10^2$
mgh10	3	$3.524 \cdot 10^9$	mgh22	20	0.4509	mgh35	10	0.8223
mgh11	3	$2.567 \cdot 10^{127}$	mgh25	10	1.282			

$\xi^1, \ldots, \xi^N$ of the random vector ξ are available, the functions a_j, b_j, and C_j, which define the surrogate functions m_j (see (1.1.3)), may be defined via the empirical averages $(1/N)\sum_{i=1}^N f_j(\cdot, \xi^i)$, $(1/N)\sum_{i=1}^N \nabla_\xi f_j(\cdot, \xi^i)$, and $(1/N)\sum_{i=1}^N \nabla_{\xi\xi} f_j(\cdot, \xi^i)$, respectively.

1.8 Supplementary Materials

1.8.1 Selection of Test Problems

For the numerical simulations in section 1.6, we selected problems from the Moré–Garbow–Hillstrom test set [237] as follows: We computed for each test problem a stationary point x_N^* of the nominal problem (1.6.3), and

$$Z_\epsilon(x_N^*) = \mathbb{E}_{\mathcal{N}(0,\epsilon I)}[X(x_N^*)] + \mathrm{StD}_{\mathcal{N}(0,\epsilon I)}[X(x_N^*)], \quad X(x_N^*)(\xi) = \frac{f_0(x_N^* + \xi) - f_0(x_N^*)}{\max\{1, |f_0(x_N^*)|\}}. \quad (1.8.1)$$

We selected those problems with $Z_\epsilon(x_N^*) \geq 1/10$ for $\epsilon = 10^{-3}$. We report the test problems and the corresponding values of $Z_\epsilon(x_N^*)$ in Table 1.6. A related approach is used by Ben-Tal and Nemirovski [25] to select uncertain linear programs from a test set.
The Moré–Garbow–Hillstrom test set is available in `Julia` through the package `NLSProblems.jl`, and in `MATLAB` through `SolvOpt` [172, 173].

1.8.2 Ambiguity Set

We show that the ambiguity set in (1.1.2) may be defined using samples of the random vector $\xi : \Omega \to \mathbb{R}^p$. We build our derivations on those established by So [300].
We provide conditions on the distribution $\mathbb{P}$ of ξ and on the sample size N such that, with high probability,

$$\begin{aligned} \|\bar{\Sigma}_N^{-1/2}(\mathbb{E}_{\mathbb{P}}[\xi] - \bar{\mu}_N)\|_2 \leq \Delta, \qquad \sigma_0\bar{\Sigma}_N \preccurlyeq \mathrm{Cov}_{\mathbb{P}}[\xi] \preccurlyeq \sigma_1\bar{\Sigma}_N \\ \ln \mathbb{E}_{\mathbb{P}}[\exp(d^T(\xi - \mathbb{E}_{\mathbb{P}}[\xi]))] \leq (1/2)\sigma_1 d^T\bar{\Sigma}_N d \quad \text{for all} \quad d \in \mathbb{R}^p. \end{aligned} \quad (1.8.2)$$

Here $\sigma_0 \leq \sigma_1$, and ξ^i, $i = 1, \ldots, N$, are independent copies of ξ. Moreover, the empirical mean and covariance matrix are defined by

$$\bar{\mu}_N = \frac{1}{N}\sum_{i=1}^N \xi^i \quad \text{and} \quad \bar{\Sigma}_N = \frac{1}{N}\sum_{i=1}^N (\xi^i - \bar{\mu}_N)(\xi^i - \bar{\mu}_N)^T, \quad (1.8.3)$$

respectively. The conditions in (1.8.2) suggest the use of the data $\bar{\mu} = \bar{\mu}_N$, $\bar{\Sigma} = \bar{\Sigma}_N$, $\bar{\Sigma}_0 = \sigma_0\bar{\Sigma}_N$, and of $\bar{\Sigma}_1 = \sigma_1\bar{\Sigma}_N$ for the definition of the ambiguity set $\mathcal{P}$ (see (1.1.2)). The first condition in (1.8.2) requires that $\bar{\Sigma}_N$ is invertible. Hence $N \geq p$.

Proposition 1.8.1. *Let $\xi : \Omega \to \mathbb{R}^p$ be a random vector with distribution $\mathbb{P} \in \mathcal{M}$ such that $\mathbb{E}_{\mathbb{P}}[\xi] \in \mathbb{R}^p$, $\mathrm{Cov}_{\mathbb{P}}[\xi] \in \mathbb{S}^p_{++}$ and*

$$\ln \mathbb{E}_{\mathbb{P}}[\exp(d^T(\xi - \mathbb{E}_{\mathbb{P}}[\xi]))] \leq (1/2) d^T \mathrm{Cov}_{\mathbb{P}}[\xi] d \quad \text{for all} \quad d \in \mathbb{R}^p. \tag{1.8.4}$$

Let $\delta \in (0, 2\mathrm{e}^{-3})$, and let ξ^i, $i = 1, \ldots, N \in \mathbb{N}$, be independent copies of ξ with

$$N > 32 \max\{(2/\mathrm{e})^2 p^2, 1\} (2\mathrm{e}/3)^3 (\ln(4p/\delta))^3. \tag{1.8.5}$$

Define $c(p) = (2/\mathrm{e})p$, and

$$\begin{aligned} t_m &= \frac{4c(p)\mathrm{e}^2(\ln(2/\delta))^2}{N}, \quad t_c = \frac{4\max\{c(p)^2, 1\}(2\mathrm{e}/3)^{3/2}(\ln(2/\delta))^{3/2}}{N^{1/2}}, \\ \Delta &= \Big(\frac{t_m}{1 - t_m - t_c}\Big)^{1/2}, \quad \sigma_0 = \frac{1}{1 + t_c}, \quad \sigma_1 = \frac{1}{1 - t_m - t_c}. \end{aligned} \tag{1.8.6}$$

Then, with probability at least $1 - \delta$, (1.8.2) *holds, where $\bar{\mu}_N$ and $\bar{\Sigma}_N$ are defined in* (1.8.3).

We apply Lemma 1.8.2 to prove Proposition 1.8.1.

Lemma 1.8.2 ([235, Lem. A.2])**.** *Let $\Sigma \in \mathbb{S}^p_{++}$ be arbitrary. Suppose that $\xi : \Omega \to \mathbb{R}^p$ is a random vector with distribution $\mathbb{P} \in \mathcal{M}$, $\mathbb{E}_{\mathbb{P}}[\xi] \in \mathbb{R}^p$, and*

$$\ln \mathbb{E}_{\mathbb{P}}[\exp(d^T(\xi - \mathbb{E}_{\mathbb{P}}[\xi]))] \leq (1/2) d^T \Sigma d \quad \text{for all} \quad d \in \mathbb{R}^p. \tag{1.8.7}$$

Then, for each $\gamma \geq 2$,

$$\begin{aligned} \mathbb{E}_{\mathbb{P}}[\|\Sigma^{-1/2}(\xi - \mathbb{E}_{\mathbb{P}}[\xi])\|_2^\gamma] &\leq 2(\gamma/\mathrm{e})^{\gamma/2} p^{\gamma/2}, \\ \mathbb{E}_{\mathbb{P}}[\|\xi - \mathbb{E}_{\mathbb{P}}[\xi]\|_2^\gamma] &\leq 2(\gamma/\mathrm{e})^{\gamma/2} (I \bullet \Sigma)^{\gamma/2}. \end{aligned} \tag{1.8.8}$$

Proof. Fix $\gamma \geq 2$. We define W, $\mathsf{Z} : \Omega \to \mathbb{R}^p$ by $\mathsf{W}(\omega) = \xi(\omega) - \mathbb{E}_{\mathbb{P}}[\xi]$ and $\mathsf{Z}(\omega) = \Sigma^{-1/2}\mathsf{W}(\omega)$, respectively. We have $\mathbb{E}_{\mathbb{P}}[\mathsf{W}] = \mathbb{E}_{\mathbb{P}}[\mathsf{Z}] = 0$. Minkowski's inequality (see, e.g., [57, p. 220]) yields

$$\mathbb{E}_{\mathbb{P}}[\|\mathsf{Z}\|_2^\gamma] \leq \Big(\sum_{i=1}^p \left(\mathbb{E}_{\mathbb{P}}\left[|\mathsf{Z}_i|^\gamma\right]\right)^{2/\gamma} \Big)^{\gamma/2} \quad \text{and} \quad \mathbb{E}_{\mathbb{P}}[\|\mathsf{W}\|_2^\gamma] \leq \Big(\sum_{i=1}^p \left(\mathbb{E}_{\mathbb{P}}\left[|\mathsf{W}_i|^\gamma\right]\right)^{2/\gamma} \Big)^{\gamma/2}. \tag{1.8.9}$$

Using (1.8.7), we obtain, for each $d \in \mathbb{R}^p$, $\ln \mathbb{E}_{\mathbb{P}}[\exp(d^T\mathsf{Z})] \leq (1/2) d^T d$. Hence, W and Z are sub-Gaussian [57, Def. 1.1 (p. 185)]. For $i = 1, \ldots, p$, we obtain from [57, Lems. 1.4 (p. 7) and 1.4 (p. 187)] that $\mathbb{E}_{\mathbb{P}}[|\mathsf{Z}_i|^\gamma] \leq 2(\gamma/\mathrm{e})^{\gamma/2}$ and $\mathbb{E}_{\mathbb{P}}[|\mathsf{W}_i|^\gamma] \leq 2(\gamma/\mathrm{e})^{\gamma/2} \Sigma_{ii}^{\gamma/2}$. Combined with (1.8.9), we deduce (1.8.8). □

Proof of Proposition 1.8.1. We apply [300, Props. 4 and 5] to establish the assertion. From (1.8.4), Lemma 1.8.2, we obtain, for each $\gamma \geq 2$,

$$\mathbb{E}_{\mathbb{P}}[\|\mathrm{Cov}_{\mathbb{P}}[\xi]^{-1/2}(\xi - \mathbb{E}_{\mathbb{P}}[\xi])\|_2^\gamma] \leq (c(p)\gamma)^{\gamma/2}. \tag{1.8.10}$$

Owing to (1.8.5) and (1.8.6), we have $t_m + t_c \in (0, 1)$ (see also [300, p. 149]). Because of (1.8.10), we can apply [300, Props. 4 and 5] to conclude that, with probability at least $1 - \delta$,

$$\|\mathrm{Cov}_{\mathbb{P}}[\xi]^{-1/2}(\mathbb{E}_{\mathbb{P}}[\xi] - \bar{\mu}_N)\|_2^2 \leq t_m, \quad (1 - t_c)\mathrm{Cov}_{\mathbb{P}}[\xi] \preccurlyeq \widehat{\Sigma}_N \preccurlyeq (1 + t_c)\mathrm{Cov}_{\mathbb{P}}[\xi], \tag{1.8.11}$$

where $\bar{\mu}_N$ is defined in (1.8.3) and $\widehat{\Sigma}_N = (1/N)\sum_{i=1}^N (\xi^i - \mathbb{E}_{\mathbb{P}}[\xi])(\xi^i - \mathbb{E}_{\mathbb{P}}[\xi])^T$ (see also [300, Thm. 9]).[3] Using (1.8.11) and the derivations in [94, p. 604], we find that

$$\widehat{\Sigma}_N = \bar{\Sigma}_N + (\bar{\mu}_N - \mathbb{E}_{\mathbb{P}}[\xi])(\bar{\mu}_N - \mathbb{E}_{\mathbb{P}}[\xi])^T \quad \text{and} \quad (\bar{\mu}_N - \mathbb{E}_{\mathbb{P}}[\xi])(\bar{\mu}_N - \mathbb{E}_{\mathbb{P}}[\xi])^T \preccurlyeq t_m \mathrm{Cov}_{\mathbb{P}}[\xi]$$

and, hence, $(1 - t_m - t_c)\mathrm{Cov}_{\mathbb{P}}[\xi] \preccurlyeq \bar{\Sigma}_N \preccurlyeq (1 + t_c)\mathrm{Cov}_{\mathbb{P}}[\xi]$. Combined with (1.8.6) and (1.8.11), we conclude that, with probability at least $1 - \delta$, $\sigma_0 \bar{\Sigma}_N \preccurlyeq \mathrm{Cov}_{\mathbb{P}}[\xi] \preccurlyeq \sigma_1 \bar{\Sigma}_N$, and, for all $d \in \mathbb{R}^p$,

$$\|\bar{\Sigma}_N^{-1/2}(\mathbb{E}_{\mathbb{P}}[\xi] - \bar{\mu}_N)\|_2^2 \leq (1 - t_m - t_c)^{-1}(\mathbb{E}_{\mathbb{P}}[\xi] - \bar{\mu}_N)^T \mathrm{Cov}_{\mathbb{P}}[\xi]^{-1}(\mathbb{E}_{\mathbb{P}}[\xi] - \bar{\mu}_N) \leq \Delta^2,$$
$$\ln \mathbb{E}_{\mathbb{P}}[\exp(d^T(\xi - \mathbb{E}_{\mathbb{P}}[\xi]))] \leq (1/2) d^T \mathrm{Cov}_{\mathbb{P}}[\xi] d \leq (1/2)\sigma_1 d^T \bar{\Sigma}_N d.$$

□

The condition in (1.8.4) holds if $\xi - \mathbb{E}_{\mathbb{P}}[\xi]$ is a strictly sub-Gaussian [57, p. 188]. For example, centered Gaussian random vectors are strictly sub-Gaussian [57, pp. 186 and 228].
The ambiguity set (1.1.2) can be defined using the data in (1.8.2) if samples of random vector ξ are available. Under the hypotheses of Proposition 1.8.1, we find that, for fixed p and δ, the scalars σ_0 and σ_1 converge to one as $N \to \infty$, and Δ converges to zero as $N \to \infty$. The convergence ensures that the data-driven ambiguity sets "shrink". However, even if $\Delta = 0$, $\sigma_0 = \sigma_1$, the ambiguity set is not a singleton in general. For example, if $p = 1$, $\bar{\Sigma}_N = 1$, $\sigma_0 = \sigma_1 = 1$, $\Delta = 0$, and $\bar{\mu}_N = 0$, then the following distributions are contained in the ambiguity set: the normal distribution $\mathcal{N}(0, 1)$, the distribution of a Rademacher random variable[4], and the uniform distribution over $[-1, 1]$ [57, pp. 13–14].

1.8.3 Formulation as Nonlinear Semidefinite Program

Proposition 1.8.3 implies that the approximated DROP (1.1.4) is equivalent to an NSDP. We recall that ψ_j is the optimal value function defined by the TRP (1.1.6), and φ_j is the one defined by the SDP (1.1.5).

Proposition 1.8.3. *Let $x \in \mathbb{R}^n$, $\rho \in \mathbb{R}$ and $j \in J$. Suppose that $\bar{\Sigma}_0 \prec \bar{\Sigma}_1$. Then $a_j(x) + \psi_j(x) + \varphi_j(x) \leq \rho$ if and only if there exists $(\gamma_j, \lambda_j, \Lambda_j, \Upsilon_j) \in \mathbb{R} \times \mathbb{R}_{\geq 0} \times \mathbb{S}_+^p \times \mathbb{S}_+^p$ such that*

$$\begin{aligned} &2a_j(x) - \gamma_j - \bar{\Sigma}_0 \bullet \Lambda_j + \bar{\Sigma}_1 \bullet \Upsilon_j \leq 2\rho, \quad \Upsilon_j - \Lambda_j = -C_j(x), \\ &\begin{bmatrix} \lambda_j I - C_j(x) & -b_j(x) \\ -b_j(x) & -\lambda_j \Delta^2 - \gamma_j \end{bmatrix} \succcurlyeq 0. \end{aligned} \tag{1.8.12}$$

We apply Lemma 1.8.4 to establish Proposition 1.8.3.

Lemma 1.8.4. *If $C \in \mathbb{S}^p$ and $\bar{\Sigma}_0 \prec \bar{\Sigma}_1$, then*

$$\min_{\Sigma \in \mathbb{S}^p} \left\{ C \bullet \Sigma : \bar{\Sigma}_0 \preccurlyeq \Sigma \preccurlyeq \bar{\Sigma}_1 \right\} = \max_{\Lambda_1, \Lambda_2 \in \mathbb{S}_+^p} \left\{ \bar{\Sigma}_0 \bullet \Lambda_1 - \bar{\Sigma}_1 \bullet \Lambda_2 : \Lambda_1 - \Lambda_2 = C \right\}, \tag{1.8.13}$$

and both problems have an optimal solution.

Proof. We transform the SDP (1.8.13) to the standard conic problem given in [26, Thm. 2.4.1 and sect. 4.1.1], and apply the duality theorem [26, Thm. 2.4.1]. We define

$$B = \begin{bmatrix} \bar{\Sigma}_0 & 0 \\ 0 & -\bar{\Sigma}_1 \end{bmatrix} \in \mathbb{S}^{2p} \quad \text{and} \quad \mathcal{A} : \mathbb{S}^p \to \mathbb{S}^{2p} \quad \text{by} \quad \mathcal{A}\Sigma = \begin{bmatrix} \Sigma & 0 \\ 0 & -\Sigma \end{bmatrix}.$$

[3] The proofs of [300, Props. 4 and 5] show that it suffices to require (1.8.10) to hold for each $\gamma \geq 2$ rather than for each $\gamma \geq 1$, as in [300, Condition (G) on p. 144].
[4] A Rademacher random variable takes the values ± 1 equiprobably.

Combined with [26, Thm. 2.4.1], we find that the optimal values in (1.8.13) is equal to

$$\min_{\Sigma\in\mathbb{S}^p} \left\{ C \bullet \Sigma : \mathcal{A}\Sigma \succcurlyeq B \right\} = \max_{\Lambda\in\mathbb{S}^{2p}_+} \left\{ B \bullet \Lambda : \mathcal{A}^*\Lambda = C \right\},$$

and these optimization problems have optimal solutions. It remains to compute the adjoint operator $\mathcal{A}^* : \mathbb{S}^{2p} \to \mathbb{S}^p$ of $\mathcal{A}$, and $B \bullet \Lambda$ for $\Lambda \in \mathbb{S}^{2p}$. We have

$$\mathcal{A}^*\Lambda = \Lambda_1 - \Lambda_2 \quad \text{with} \quad \Lambda = \begin{bmatrix} \Lambda_1 & \Lambda_3 \\ \Lambda_3^T & \Lambda_2 \end{bmatrix} \in \mathbb{S}^{2p}.$$

Indeed, for each $\Sigma \in \mathbb{S}^p$ and $\Lambda \in \mathbb{S}^{2p}$, we have $(\mathcal{A}\Sigma) \bullet \Lambda = \Sigma\Lambda_1 - \Sigma\Lambda_2 = \Sigma \bullet (\mathcal{A}^*\Lambda)$. Moreover, for each $\Lambda \in \mathbb{S}^{2p}$, $B \bullet \Lambda = \bar{\Sigma}_0 \bullet \Lambda_1 - \bar{\Sigma}_1 \bullet \Lambda_2$. □

Proof of Proposition 1.8.3. Using the strong duality for the TRP (1.1.6) [53, sect. B.1, eq. (B.2)] (see also [306, Cor. 5.3]), we find that

$$\begin{aligned}
2\psi_j(x) &= \max_{s\in\mathbb{R}^p} \left\{ 2b_j(x)^T s + s^T C_j(x) s : \|s\|_2^2 \le \Delta^2 \right\} \\
&= -\min_{s\in\mathbb{R}^p} \left\{ -2b_j(x)^T s - s^T C_j(x) s : \|s\|_2^2 \le \Delta^2 \right\} \\
&= -\max_{(\gamma_j,\lambda_j)\in\mathbb{R}^2} \left\{ \gamma_j : \begin{bmatrix} \lambda_j I - C_j(x) & -b_j(x) \\ -b_j(x) & -\lambda_j\Delta^2 - \gamma_j \end{bmatrix} \succcurlyeq 0, \quad \lambda_j \ge 0 \right\} \\
&= \min_{(\gamma_j,\lambda_j)\in\mathbb{R}^2} \left\{ -\gamma_j : \begin{bmatrix} \lambda_j I - C_j(x) & -b_j(x) \\ -b_j(x) & -\lambda_j\Delta^2 - \gamma_j \end{bmatrix} \succcurlyeq 0, \quad \lambda_j \ge 0 \right\}.
\end{aligned}$$

Lemma 1.8.4 and the definition of φ_j provided in (1.1.5) give

$$\begin{aligned}
2\varphi_j(x) &= \max_{\Sigma\in\mathbb{S}^p} \left\{ C_j(x) \bullet \Sigma : \bar{\Sigma}_0 \preccurlyeq \Sigma \preccurlyeq \bar{\Sigma}_1 \right\} \\
&= \min_{\Lambda_j,\,\Upsilon_j\in\mathbb{S}^p_+} \left\{ -\bar{\Sigma}_0 \bullet \Lambda_j + \bar{\Sigma}_1 \bullet \Upsilon_j : \Upsilon_j - \Lambda_j = -C_j(x) \right\}.
\end{aligned}$$

Now, we use the arguments from the proof of [23, Thm. 2.1] to deduce the equivalence of $a_j(x) + \psi_j(x) + \varphi_j(x) \le \rho$ and (1.8.12). □

1.8.4 Performance of SDP Solvers on Box-Constrained SDPs

We compare the performance of state-of-the-art SDP solvers applied to a sequence of the "(Löwner-)box-constrained" SDPs

$$\max_{X\in\mathbb{S}^p} C \bullet X \quad \text{s.t.} \quad 0 \preccurlyeq X \preccurlyeq I.$$

with an implementation of the formulas for the optimal value and an optimal solution provided by Proposition 1.3.1. The performance is compared in terms of run time, number of allocations and memory usage. Box-constrained SDPs also appear in [23, sect. 4.5.6 and eq. (4.5.45)], where tractable approximations of quadratically perturbed chance constraints are derived.

We defined the matrices $C \in \mathbb{S}^p$ for $p \in \{5, 10, 15, \ldots, 100\}$ using the symmetric part of randomly generated matrices. The entries of these random matrices are realizations of independent standard normal random variables. For each $p \in \{5, 10, 15, \ldots, 100\}$, we generated ten independent realizations of $C \in \mathbb{S}^p$ which define a sequence of box-constrained SDPs.

We implemented the solution formulas provided by Proposition 1.3.1 in `Julia` [35] (version 1.4.2) without preallocated memory for the intermediate calculations, such as matrix multiplications, and call the implementation `EigSDP`. The box-constrained SDPs were modeled in `Julia` using

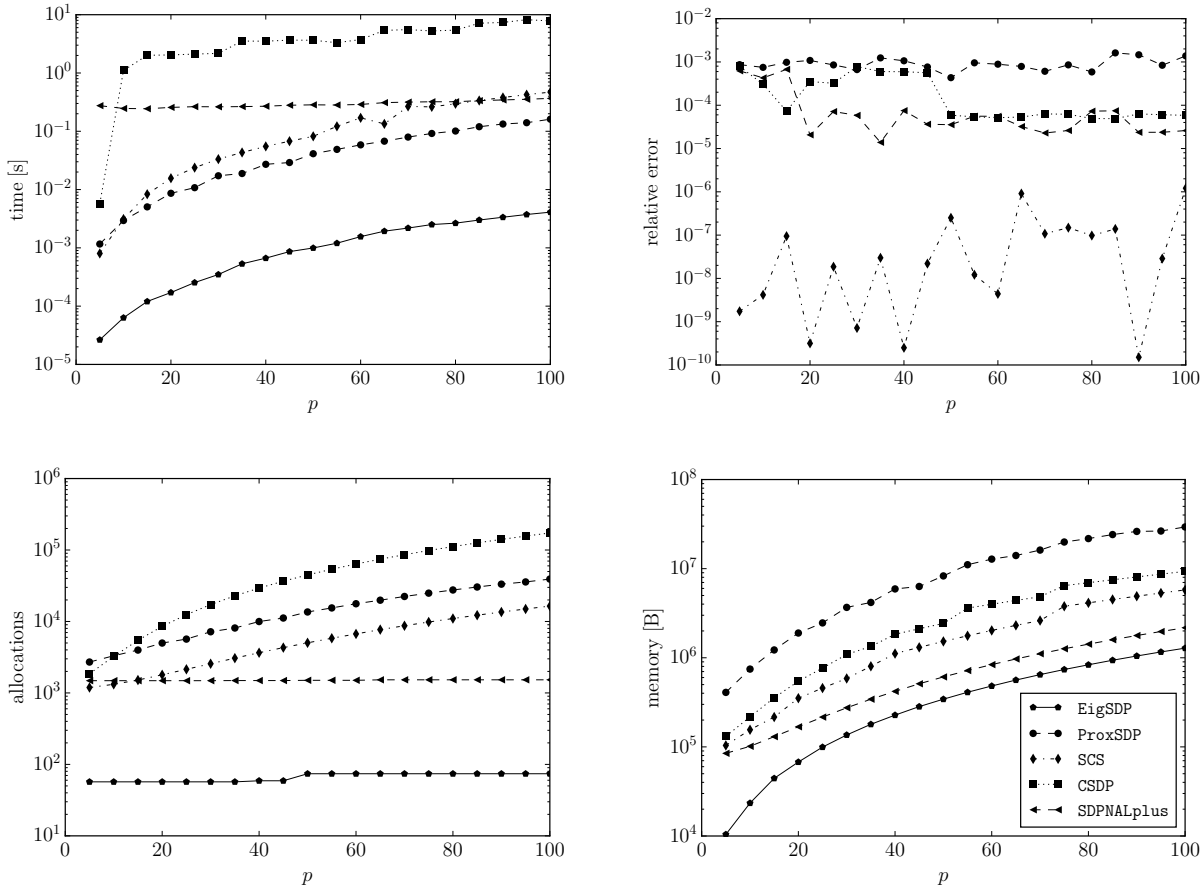


Figure 1.1: *Median of solution times (in seconds), relative errors objective function values, allocations, and memory usage (in bytes) computed over ten independently generated box-constrained SDPs. The legend applies to all subfigures.*

`JuMP` [102]. `JuMP` allows us to efficiently model and solve a sequence of the box-constrained SDPs without parsing each problem again when the solution command is called. For each of the ten independent realizations, we used the `JuMP`-macro `@objective` to modify the objective function of the `JuMP` models. We used the SDP solvers `SCS` [251, 252], `SDPNALplus` [307, 353, 360], `CSDP` [47, 48], and `ProxSDP.jl` [302]. For the first three codes, we utilized the `Julia` wrappers `SCS.jl`, `SDPNAL.jl`, and `CSDP.jl`. The interface `SDPNAL.jl` was used in combination with `MATLAB` (version R2020a) and `MATLAB.jl`.

The tolerances for the primal and dual feasibility, and the relative duality gap were set to 10^{-3} as in [251, sect. 6.1]. For the remaining settings, we used the solvers' defaults (versions as of May 23, 2020). We measured the elapsed time and the total bytes allocated with the `Julia` macro `@timed`, and computed the allocations with `@timed` and `gc_alloc_count`, which is part of `Julia`'s `Base`-module.

The comparison was made on a computer with an Intel Core i5-4590T processor with 2 GHz, and 16 GB of RAM. Figure 1.1 depicts the median of the elapsed solution time in seconds, of the allocations, of the memory usage in bytes, and of the relative error of the objective function values. To compute the relative errors, we used the optimal values returned by `EigSDP` as a reference. The solution time and the number of allocations of `EigSDP` is significantly lower than those of the SDP solvers. We observe that `SCS` solves the SDPs more accurately than required by our termination tolerance.

2 Approximation Scheme for Distributionally Robust PDE-Constrained Optimization

We extend the sampling-free approximation scheme and the algorithmic approach developed in Chapter 1 to distributionally robust optimization problems (DROPs) with parameterized partial differential equations (PDEs). We prove the existence of optimal solutions for the DROP, and of the approximated and smoothed DROPs, and show that a worst-case distribution exists. Moreover, we analyze the approximation error resulting from the second-order Taylor's expansion. The smoothing method provided in Algorithm 1 is extended to allow the numerical treatment of DROPs posed in Hilbert spaces using gradient-based optimization methods. The adjoint approach is used to compute derivatives of the smoothing functions. We present numerical results for the distributionally robust optimization of the steady and of the unsteady Burgers' equations.
The chapter is mainly based on the manuscript

[235] J. Milz and M. Ulbrich, *An approximation scheme for distributionally robust PDE-constrained optimization*, Preprint No. IGDK-2020-09. Technische Universität München, München, Jun. 2020, in review, `http://www.igdk.eu/foswiki/pub/IGDK1754/Preprints/MilzUlbrich-PDEDRO.pdf`.

Section 2.3 provides further specifics on the existence of optimal solutions than [235, sect. 3], and section 2.7 contains more details on the analysis of the control problems than that in [235, sect. 7]. We present a derivation of the derivatives required by our approximation scheme using the adjoint approach in section 2.9.3.

2.1 Introduction

We consider the distributionally robust nonlinear, PDE-constrained optimal control problem

$$\min_{u\in U_{\mathrm{ad}}} \Big\{ \sup_{P\in\mathcal{P}} \mathbb{E}_P[J(S(u,\xi),u,\xi)] \Big\}, \tag{2.1.1}$$

where $\mathbb{E}_P[J(S(u,\xi),u,\xi)] = \int_{\mathbb{R}^p} J(S(u,\xi),u,\xi)\mathrm{d}P(\xi)$, $U_{\mathrm{ad}} \subset U$ is the set of admissible controls, and U is a Hilbert space. Moreover, $J : Y \times U \times \mathbb{R}^p \to \mathbb{R}$ is the *parametrized objective function*, and $S : U \times \mathbb{R}^p \to Y$ is the *solution operator* of the parameterized PDE $e(S(u,\xi),u,\xi) = 0$ for $\xi \in \mathbb{R}^p$, where $e : Y \times U \times \mathbb{R}^p \to Z$, and Y, Z are Banach spaces. Throughout the chapter, the ambiguity set $\mathcal{P}$ is defined by

$$\begin{aligned}\mathcal{P} = \{P \in \mathcal{M} :\ & \|\bar{\Sigma}^{-1/2}(\mathbb{E}_P[\xi] - \bar{\mu})\|_2 \le \Delta, \quad \sigma_0\bar{\Sigma} \preccurlyeq \mathrm{Cov}_P[\xi] \preccurlyeq \sigma_1\bar{\Sigma}, \\ & \ln \mathbb{E}_P[\exp\left(d^T(\xi - \mathbb{E}_P[\xi])\right)] \le (\sigma_1/2)d^T\bar{\Sigma}d \quad \text{for all} \quad d \in \mathbb{R}^p \},\end{aligned} \tag{2.1.2}$$

where $\Delta > 0$, $\bar{\mu} \in \mathbb{R}^p$, $\bar{\Sigma} \in \mathbb{S}^p_{++}$, and σ_0, $\sigma_1 \in \mathbb{R}_+$ fulfill $\sigma_0 < \sigma_1$. The ambiguity set $\mathcal{P}$ is an instance of that given in (1.1.2). Its definition is motivated by the results established in section 1.8.2.

We define the *reduced parametrized objective function* $\widehat{J} : U \times \mathbb{R}^p \to \mathbb{R}$ associated with (2.1.1) by

$$\widehat{J}(u, \xi) = J(S(u, \xi), u, \xi). \tag{2.1.3}$$

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

$$Q(u, \xi; \bar{\mu}) = \widehat{J}(u, \bar{\mu}) + \nabla_\xi \widehat{J}(u, \bar{\mu})^T (\xi - \bar{\mu}) + (1/2)(\xi - \bar{\mu})^T \nabla_{\xi\xi} \widehat{J}(u, \bar{\mu})(\xi - \bar{\mu}), \tag{2.1.4}$$

where $Q(\cdot, \cdot; \bar{\mu}) : U \times \mathbb{R}^p \to \mathbb{R}$. Since $Q(u, \cdot; \bar{\mu})$ is quadratic, the expected value $\mathbb{E}_P[Q(u, \xi; \bar{\mu})]$ can be computed explicitly for $u \in U$; see section 1.1.
As in section 1.1, we approximate the DROP (2.1.1) with

$$\min_{u \in U_{\mathrm{ad}}} \Big\{ \sup_{P \in \mathcal{P}} \mathbb{E}_P[Q(u, \xi; \bar{\mu})] \Big\}. \tag{2.1.5}$$

We refer to the control problem (2.1.5) as approximated DROP. From section 1.1, we obtain that the objective function $F : U_{\mathrm{ad}} \to \mathbb{R}$ of (2.1.5) can be written as

$$F(u) = \widehat{J}(u, \bar{\mu}) + \varphi(u) + \psi(u), \tag{2.1.6}$$

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

$$\varphi(u) = \max_{\Sigma \in \mathbb{S}^p} \left\{ (1/2) \nabla_{\xi\xi} \widehat{J}(u; \bar{\mu}) \bullet \Sigma : \ \sigma_0 \bar{\Sigma} \preccurlyeq \Sigma \preccurlyeq \sigma_1 \bar{\Sigma} \right\} \tag{2.1.7}$$

and $\psi : U \to \mathbb{R}$ is the optimal value function of nonconvex TRP

$$\psi(u) = \max_{d \in \mathbb{R}^p} \left\{ \nabla_\xi \widehat{J}(u; \bar{\mu})^T d + (1/2) d^T \nabla_{\xi\xi} \widehat{J}(u; \bar{\mu}) d : \ \|\bar{\Sigma}^{-1/2} d\|_2 \leq \Delta \right\}. \tag{2.1.8}$$

Under suitable assumption, we show that the approximation error of the objective function of (2.1.1) and that of (2.1.5) is small. The cost function F of (2.1.5) can efficiently be evaluated (being the sum of a TRP and an SDP) without further approximations, such as sampling; however it is nonsmooth; see section 1.1. For the numerical solution of (2.1.5), we construct a smoothing function for F which we use to define smoothed DROPs similar to those in section 1.2. We extend the smoothing method developed in section 1.2 to allow the numerical treatment of infinite-dimensional problems. All required derivatives of the smoothing function for F are computed using `UFL` [7, 5] and `FEniCS` [6, 220]. Furthermore, we prove the existence of optimal solutions for the DROP (2.1.1), of the approximated DROP (2.1.5) and smoothed DROPs defined in (2.2.1). In addition, we establish the existence of a worst-case distribution of (2.1.1). Moreover, we present a convergence result of the smoothing method.

Related Work

A popular solution approach (see, e.g., [94, 344, 75]) for moment-based DROPs exploits the fact that a Lagrangian dual of the maximization problem in (2.1.1) is a robust optimization problem [94, sect. 2.1]. Under suitable assumptions, strong duality holds and the dual can be concatenated with the upper-level problem to obtain a single-level problem [94, 344]. The tractability of the dual depends on the structure of $\widehat{J}(u, \cdot)$ [94, 344, 23]. For example, if $\widehat{J}(u, \cdot)$ is the pointwise maximum of affine functions for all $u \in U$, the dual is tractable [94, sect. 4.1]. However, this approach does not result in explicit single-level programs when $\widehat{J}(u, \cdot)$ is

implicitly defined or non-quadratic. Moreover, a "robustified" constraint appears in the single-level problem [94, p. 597] and, hence, available solvers for PDE-constrained problems cannot be applied to it.

The DROP (2.1.1) becomes a risk-neutral problem when the ambiguity set $\mathcal{P}$ is a singleton. We refer the reader to [294] for an overview of stochastic programming. Risk-neutral optimization with PDEs has been considered, for example, in [123, 189].

The objective function $\widehat{f} : U \to \mathbb{R} \cup \{\infty\}$ of (2.1.1),

$$\widehat{f}(u) = \sup_{P \in \mathcal{P}} \mathbb{E}_P[\widehat{J}(u, \xi)], \tag{2.1.9}$$

is convex if $\widehat{J}(\cdot, \xi)$ is convex for all $\xi \in \mathbb{R}^p$ and if $\mathbb{E}_P[\widehat{J}(u, \xi)]$ is well-defined for all $P \in \mathcal{P}$. Similarly, the well-defined composition of a convex, monotonic risk measure with $u \mapsto \widehat{J}(u, \cdot)$ is convex if $\widehat{J}(\cdot, \xi)$ is convex for all $\xi \in \mathbb{R}^p$ [294, Prop. 6.11].

Risk-averse optimization problems with coherent risk measures can equivalently be reformulated as min-max problems similar to those in (2.1.1) where the ambiguity set is the domain of the convex conjugate of the risk measure [280, 294]. In the literature on optimization with PDEs, popular risk measures are: the superquantile/conditional value-at-risk [190, 191], the entropic risk measure [192], and the mean-plus-variance risk measure [2, 71, 325, 29, 230]. Risk-neutral and risk-averse shape optimization are investigated in [84, 85, 285]. Existence results for solutions and optimality conditions for risk-averse control problems with PDEs are provided in [190, 191]. Optimality conditions are derived in [130] for risk-neutral convex optimization problems posed in Banach spaces with almost sure state constraints. The entropic risk measure, also called log-exponential risk measure [269], was introduced by Whittle [342], [343, sect. 19] in the context of stochastic control.

DRO with PDEs is considered by Kouri [186] with ambiguity sets including the support constraint $P(\xi \in \Xi) = 1$, where $\Xi \subset \mathbb{R}^p$ is a convex, compact Lipschitz domain. An inner approximation of the ambiguity set is constructed via a measure discretization, error bounds are derived and Kouri [186] has shown that the objective function of the min-max problem is Clarke-subdifferentiable. Even though the cost function is subdifferentiable, only a limited number of algorithms for nonsmooth control problems in infinite dimensional spaces are available; we refer the reader to [147] and the references therein. Our approach allows us to apply gradient-based solvers for PDE-constrained optimization problems. The ambiguity set (2.1.2) does not have a support constraint and, hence, it violates the assumptions made in [186]. For the approach developed in [186], the number of solutions of the parameterized state equation depends on the measure discretization. Our scheme is sampling-free and the evaluation of the smoothing function for (2.1.6) requires only one solution of the state equation.

Taylor's expansions have also been used to approximate mean-plus-variance minimization problems with PDEs [2, 71] and robust nonlinear optimization problems with PDEs [4, 90, 181, 209]. The authors of [2] and of [71] develop an approximation scheme for mean-plus-variance minimization problems with PDEs depending on a random field. The parameterized cost function is approximated using first- and second-order Taylor's expansions, allowing the authors to explicitly compute the expectation and variance of the surrogate objective function. The objective function of a mean-plus-variance minimization problem may be nonconvex even if $\widehat{J}(\cdot, \xi)$ is convex for each $\xi \in \mathbb{R}^p$ [296, p. 114].

If the ambiguity set $\mathcal{P}$ is given by all probability distributions supported on a (compact) set Ξ, that is, if $\mathcal{P} = \{ P \in \mathcal{M} : P(\Xi) = 1 \}$, then the DROP (2.1.1) is equivalent to a robust optimization problem [295, p. 535]. Robust optimization with PDEs is considered, for example, in [4, 90, 181, 209, 148, 298]. For numerical computations, the authors of [4, 181, 209] use second-order Taylor's expansions and obtain (2.1.5) without the optimal value function defined

Algorithm 2 Smoothing method

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

1. Compute a stationary point u_k of (2.2.1) using u_{k-1} as initial point.
2. Choose $0 < (\tau_{k+1}, \nu_{k+1}, \eta_{k+1}) < (\tau_k, \nu_k, \eta_k)$.

by the SDP (2.1.7), and they either reformulate the constraints given by the TRP (2.1.8) using its necessary and sufficient optimality conditions, or use smooth optimization methods. The first approach results in a mathematical problem with complementarity problems and linear matrix inequalities [209, sect. 3.2.2]. Depending on the size of the trust-region radius in (2.1.8), several optimal solutions of (2.1.8) may exist [238, p. 556] and, hence, the optimal value function (2.1.8) may be nondifferentiable. Optimization methods for smooth problems may therefore not be a suitable class of algorithms. Our algorithmic scheme provides an alternative to those used in [4, 181, 209].
Smoothing schemes provide a popular algorithmic approach for nonsmooth PDE-constrained optimization problems. For example, Mannel (*né* Kruse) and Ulbrich [198] develop an interior-point approach for optimal control with state constraints using a smoothed constraint function, Kouri and Surowiec [191, 193] propose smoothing schemes for risk-averse PDE-constrained optimization, and an interior-point approach for risk-averse PDE-constrained optimization is developed in [122].
For an overview of recent contributions to and challenges of the field of PDE-constrained optimization, we refer the reader to [322, 80].

2.2 Smoothing Functions and Smoothing Method

We describe our algorithmic approach, which is based on that developed in section 1.2, to compute a stationary point of the approximated DROP (2.1.5). We construct smoothing functions $\widetilde{\psi} : U \times \mathbb{R}^2_{++} \to \mathbb{R}$ of $\psi : U \to \mathbb{R}$ (see (2.1.8)) and $\widetilde{\varphi} : U \times \mathbb{R}_{+} \to \mathbb{R}$ of $\varphi : U \to \mathbb{R}$ (see (2.1.7)), and compute approximate stationary points of a sequence of smoothed control problems

$$\min_{u \in U_{\mathrm{ad}}} \left\{ \widetilde{F}(u; \tau_k, \nu_k, \eta_k) = \widehat{J}(u, \bar{\mu}) + \widetilde{\varphi}(u; \tau_k) + \widetilde{\psi}(u; \nu_k, \eta_k) \right\} \tag{2.2.1}$$

with decreasing smoothing parameters $(\tau_k, \nu_k, \eta_k) \in \mathbb{R}^3_{++}$ indexed by the outer iteration counter k, where $\widetilde{F} : U \times \mathbb{R}^3_{++} \to \mathbb{R}$. We summarize this algorithmic framework in Algorithm 2. In section 2.4, we prove the convergence of weak limit points of optimal solutions, generated by Algorithm 2, to minimizers of (2.1.5). We can apply the same gradient-based optimization methods to (2.2.1) that are suitable for the nominal control problem

$$\min_{u \in U_{\mathrm{ad}}} \widehat{J}(u, \bar{\mu}). \tag{2.2.2}$$

We extend the notion of smoothing functions provided in Definition 1.2.1 to functions defined on (infinite-dimensional) Banach spaces. Definition 2.2.1 is also based on that in [73, Def. 3.1].

Definition 2.2.1. *Let X be a Banach space and let $\phi : X \to \mathbb{R}$ be continuous. A function $\widetilde{\phi} : X \times \mathbb{R}^m_{++} \to \mathbb{R}$ is a* smoothing function *for ϕ if $\widetilde{\phi}(\cdot; t)$ is continuously differentiable for all $t \in \mathbb{R}^m_{++}$, and there exists $\gamma : \mathbb{R}^m_{+} \to \mathbb{R}_{+}$ with $\gamma(t) \to 0$ as $\mathbb{R}^m_{++} \ni t \to 0$, such that, for each $x \in X$ and $t > 0$, we have $|\phi(x) - \widetilde{\phi}(x; t)| \leq \gamma(t)$.*

2.2.1 Smoothing Approach for the SDP

Based on the construction made in section 1.3, we state a smoothing function for the optimal value function φ defined in (2.1.7) and show that is satisfies the conditions of Definition 2.2.1. Throughout the section, let $\nabla_{\xi\xi}\widehat{J}(\cdot,\bar{\mu})$ be continuously differentiable. Using Proposition 1.3.1, we have

$$\varphi(u) = (\sigma_0/2)G(u) \bullet I + ((\sigma_1 - \sigma_0)/2)\sum_{i=1}^{p}(\lambda_i\left(G(u)\right))_+, \tag{2.2.3}$$

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

$$G(u) = \bar{\Sigma}^{1/2}\nabla_{\xi\xi}\widehat{J}(u,\bar{\mu})\bar{\Sigma}^{1/2}. \tag{2.2.4}$$

Using (2.2.3), and the continuity of λ [157, Cor. 6.3.8] and of $\nabla_{\xi\xi}\widehat{J}(\cdot,\bar{\mu})$, we find that φ_j is continuous. We define $\widetilde{\varphi} : U \times \mathbb{R}_{++} \to \mathbb{R}$ by

$$\widetilde{\varphi}(u;\tau) = (\sigma_0/2)G(u) \bullet I + ((\sigma_1 - \sigma_0)/2)\widetilde{w}(\lambda(G(u));\tau), \tag{2.2.5}$$

where $\widetilde{w} : U \times \mathbb{R}_{++} \to \mathbb{R}$ is defined in (1.3.3). We show that $\widetilde{\varphi}$ is a smoothing function for φ. We fix $\tau > 0$. Using (1.3.5), we find that

$$\varphi(u) \leq \widetilde{\varphi}(u;\tau) \leq \varphi(u) + (1/2)\tau p \ln 2 \quad \text{for all} \quad u \in U. \tag{2.2.6}$$

The mapping $\mathbb{S}^p \ni A \mapsto \widetilde{w}(\lambda(A);\tau)$ is twice continuously differentiable [217, Thm. 4.2]. Combined with (2.2.6), we conclude that $\widetilde{\varphi}$ is a smoothing function for φ. Moreover, if $\nabla_{\xi\xi}\widehat{J}(\cdot,\bar{\mu})$ is twice continuously differentiable, then $\widetilde{\varphi}(\cdot;\tau)$ is twice continuously differentiable.

2.2.2 Smoothing Approach for the TRP

Based on the construction made in section 1.4, we state a smoothing function of the optimal value function φ defined in (2.1.8) and show that it satisfies the conditions of Definition 2.2.1. Throughout the section, let $\nabla_{\xi}\widehat{J}(\cdot,\bar{\mu})$ and $\nabla_{\xi\xi}\widehat{J}(\cdot,\bar{\mu})$ be continuously differentiable. We define $\widetilde{\psi} : U \times \mathbb{R}^2_{++} \to \mathbb{R}$ by

$$\widetilde{\psi}(u;\nu,\eta) = \max_{\widetilde{s}\in\mathbb{R}^{p+2}} \left\{ \widetilde{g}_\nu(u)^T\widetilde{s} + (1/2)\widetilde{s}^T\widetilde{H}_\eta(u)\widetilde{s} : \ \ (1/2)\|\widetilde{s}\|_2^2 \leq (1/2)\Delta^2 \right\}, \tag{2.2.7}$$

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

$$\widetilde{H}_\eta(u) = \begin{bmatrix} G(u) & & \\ & 0 & \\ & & E(G(u);\eta) \end{bmatrix} \quad \text{and} \quad \widetilde{g}_\nu(u) = \begin{bmatrix} g(u) \\ \sqrt{2\nu} \\ \sqrt{2\nu} \end{bmatrix}. \tag{2.2.8}$$

Moreover, ν, $\eta > 0$, and the "preconditioned" gradient mapping $g : U \to \mathbb{R}^p$ is given by

$$g(u) = \bar{\Sigma}^{1/2}\nabla_{\xi}\widehat{J}(u,\bar{\mu}). \tag{2.2.9}$$

Here, $E : \mathbb{S}^p \times \mathbb{R}_{++} \to \mathbb{R}$ is the entropy function defined in (1.4.10), which is twice continuously differentiable [217, Thm. 4.2], and G is defined in (2.2.4).
Using (2.1.8) and [154, Thm. 7], we find that φ is continuous. Let ν, $\eta > 0$ be arbitrary. From (1.4.31), we obtain

$$\psi(u) \leq \widetilde{\psi}(u;\nu,\eta) \leq \psi(u) + 2\sqrt{2\nu}\Delta + (1/2)\Delta^2\eta \ln p \quad \text{for all} \quad u \in U. \tag{2.2.10}$$

We establish the continuous differentiability of $\widetilde{\psi}(\cdot;\nu,\eta)$ without relying on Lagrangian duality as in the proof of Theorem 1.4.9.
We show that an optimal solution to (2.2.7) is unique. Fix $u \in U$ and let $(\widetilde{s},\widetilde{\lambda})$ be a KKT-tuple of (2.2.7). Using Theorem 1.4.1, we obtain $\widetilde{\lambda} \geq 0$, $(\widetilde{H}_\eta(u) - \widetilde{\lambda}I)\widetilde{s} = \widetilde{g}_\nu(u)$, $\widetilde{H}_\eta(u) - \widetilde{\lambda}I \preccurlyeq 0$, $\widetilde{\lambda}(\|\widetilde{s}\|_2^2 - \Delta^2) = 0$ and, moreover, if $\lambda_{\max}(\widetilde{H}_\eta(u)) < \widetilde{\lambda}$, then $\widetilde{s}$ is the unique solution to (2.2.7). Using (2.2.8), we have $\widetilde{\lambda} > (E(G(u);\eta))_+$; otherwise the linear system $(\widetilde{H}_\eta(u) - \widetilde{\lambda}I)\widetilde{s} = \widetilde{g}_\nu(u)$ would not have a solution. It remains to compute $\lambda_{\max}(G(u))$. According to (1.4.11), we have $\lambda_{\max}(G(u)) \leq E(G(u);\eta)$. Combined with (2.2.8), we obtain $\lambda_{\max}(\widetilde{H}_\eta(u)) = (E(G(u);\eta))_+$. Putting together the pieces, we find that, for each $u \in U$, the TRP (2.2.7) has a unique solution. Danskin's theorem [46, Thm. 4.13 and Rem. 4.14] implies that $\widetilde{\psi}(\cdot;\nu,\eta)$ is differentiable. Moreover, the optimal solution to (2.2.7) as a function of the control is continuous [154, Cor. 8.2]. Hence $\widetilde{\psi}(\cdot;\nu,\eta)$ is continuously differentiable. Putting together the statements, we conclude that $\widetilde{\psi}$ is a smoothing function for φ
If $\nabla_\xi \widehat{J}(\cdot,\bar{\mu})$ and $\nabla_{\xi\xi}\widehat{J}(\cdot,\bar{\mu})$ are twice continuously differentiable, then the implicit function theorem, applied to the above first-order optimality conditions of (2.2.7), implies that the optimal solution to (2.2.7) as a function of the control is continuously differentiable. The chain rule implies that $\widetilde{\psi}(\cdot;\nu,\eta)$ is twice continuously differentiable.

2.3 Existence of Optimal Solutions

We prove the existence of optimal solutions of the DROP (2.1.1) and of the maximization problem in (2.1.1). We refer to an optimal solution of the maximization problem in (2.1.1) as a *worst-case distribution*. Moreover, we prove that there exist minimizers of the approximated DROP (2.1.5) and of the smoothed DROPs (2.2.1).

2.3.1 Existence of Optimal Solutions of the DROP

We state conditions implying that (2.1.1) has optimal solutions which are built on those used by Kouri and Shapiro [190, Chap. 3] and by Kouri and Surowiec [192, sect. 3.2].

Assumption 2.3.1 ([235, Assumption 3.1])**.** *For each $(u,\xi) \in U_{ad} \times \mathbb{R}^p$, let $S(u,\xi) \in Y$ be the unique solution to: Find $y \in Y$: $e(y,u,\xi) = 0$, where $S : U_{ad} \times \mathbb{R}^p \to Y$ and $e : Y \times U \times \mathbb{R}^p \to Z$.*
(a) *For every $\xi \in \mathbb{R}^p$, $S(\cdot,\xi) : U_{ad} \to Y$ is weakly-weakly continuous.*
(b) *For all $u \in U_{ad}$, $S(u,\cdot) : \mathbb{R}^p \to Y$ is continuous.*

Assumption 2.3.1 implies that the reduced parameterized objective function $\widehat{J}$ defined in (2.1.3) is well-defined. Assumption 2.3.1 (a) may be verified using Proposition 2.9.3. If $e_y(S(u,\xi),u,\xi) \in \mathscr{L}(Y,Z)$ is boundedly invertible for each $(u,\xi) \in U_{\mathrm{ad}} \times \mathbb{R}^p$, then the implicit function theorem implies Assumption 2.3.1 (b).

Assumption 2.3.2 ([235, Assumption 3.2])**.** *The function $J : Y \times U_{ad} \times \mathbb{R}^p \to \mathbb{R}$ is continuous.*
(a) *There exists $\gamma \in \mathbb{R}$ such that $J(y,u,\xi) \geq \gamma$ for all $(y,u,\xi) \in Y \times U_{ad} \times \mathbb{R}^p$.*
(b) *For each $\xi \in \mathbb{R}^p$, $J(\cdot,\cdot,\xi) : Y \times U_{ad} \to \mathbb{R}$ is weakly lower semicontinuous.*

We show in Lemma 2.3.5 that Assumption 2.3.2 is satisfied for tracking-type functionals. Assumption 2.3.2 (b) holds if $J(\cdot,\cdot,\xi)$ is convex for all $\xi \in \mathbb{R}^p$ and J is continuous [151, Thm. 1.18], [46, pp. 26–27]. Assumptions 2.3.1 and 2.3.2 imply that $\widehat{J}(u,\cdot) = J(S(u,\cdot),u,\cdot)$ is continuous for each fixed $u \in U_{\mathrm{ad}}$. Hence $\widehat{J}(u,\cdot)$ is measurable for all $u \in U_{\mathrm{ad}}$ [169, Lem. 1.5].
We introduce the notion of uniform integrability. A measurable function $\theta : \mathbb{R}^p \to \mathbb{R}$ is uniformly integrable (w.r.t. $\mathcal{P}$) if $\sup_{P\in\mathcal{P}} \mathbb{E}_P[|\theta(\xi)|1_{|\theta(\xi)|\geq t}] \to 0$ as $t \to \infty$ [45, sect. 2.7(i)]. Throughout, uniform integrability is meant w.r.t. the probability measures contained in the ambiguity set $\mathcal{P}$ defined in (2.1.2).

Assumption 2.3.3 ([235, Assumption 3.3]). *For each $u \in U_{ad}$, the function $\widehat{J}(u,\cdot)$ defined in* (2.1.3) *is uniformly integrable.*

Assumption 2.3.3 can be verified for many control problems using Lemmas 2.3.4 and 2.3.5.

Lemma 2.3.4 ([235, Lem. 3.4]). *The following conditions ensure the uniform integrability of the measurable function $h : \mathbb{R}^p \to \mathbb{R}$:* (a) $\sup_{P\in\mathcal{P}} \mathbb{E}_P[|h(\xi)|^r] < \infty$ *for some $r > 1$;* (b) *there exists a uniformly integrable function $h_1 : \mathbb{R}^p \to \mathbb{R}_+$ such that $|h| \le h_1$;* (c) *h_1, $h_2 : \mathbb{R}^p \to \mathbb{R}$ are uniformly integrable and $h = h_1 + h_2$.*

Proof. Let $t > 0$ and $P \in \mathcal{P}$ be arbitrary. The first assertion follows from $\mathbb{E}_P[|h(\xi)|1_{|h(\xi)|>t}] \le t^{-r+1}\mathbb{E}_P[|h(\xi)|^r]$ (see, e.g., [169, p. 67]), and the second statement follows from the fact that h_1 dominates $|h|$. The estimate $\mathbb{E}_P[|h(\xi)|1_{|h(\xi)|\ge 2t}] \le 2\mathbb{E}_P[|h_1(\xi)|1_{|h_1(\xi)|\ge t}] + 2\mathbb{E}_P[|h_2(\xi)|1_{|h_2(\xi)|\ge t}]$ (see, e.g., [39, p. 230]) implies the third claim. □

Lemma 2.3.5 is built on [235, Ex. 3.5].

Lemma 2.3.5. *Let H be a Hilbert space, and let $Y \hookrightarrow H$ be a continuous embedding. Let $y_d \in H$ and $\alpha \ge 0$. We define the tracking-type function $J : Y \times U \times \mathbb{R}^p \to \mathbb{R}$ by $J(y,u,\xi) = (1/2)\|y - y_d\|_H^2 + (\alpha/2)\|u\|_U^2$. If Assumption 2.3.1 holds, then Assumption 2.3.2 is satisfied. If, in addition, $u \in U$ and $\|S(u,\cdot)\|_H^2$ is uniformly integrable, then the function $\widehat{J}(u,\cdot)$ defined in* (2.1.3) *is uniformly integrable.*

Proof. To verify Assumption 2.3.2, we first observe that J is independent of ξ and $J \ge 0$. Since $J(\cdot,\cdot,\xi)$ is convex and continuous for each $\xi \in \mathbb{R}^p$, Assumption 2.3.2 (b) holds [151, Thm. 1.18]. It must yet be shown that the uniform integrability of $\|S(u,\cdot)\|_H^2$ implies that of $\widehat{J}(u,\cdot)$. Using Young's inequality, we have, for each $\xi \in \mathbb{R}^p$,

$$\widehat{J}(u,\xi) = \tfrac{1}{2}\|S(u,\xi) - y_d\|_H^2 + \tfrac{\alpha}{2}\|u\|_U^2 \le \|S(u,\xi)\|_H^2 + \|y_d\|_H^2 + \tfrac{\alpha}{2}\|u\|_U^2.$$

Now, Lemma 2.3.4 implies that $\widehat{J}(u,\cdot)$ is uniformly integrable. □

We verify the uniform integrability of $\|S(u,\cdot)\|_H^2$ for two parameterized Burgers' equations in section 2.7. Next, we show that the DROP (2.1.1) has an optimal solution.

Theorem 2.3.6 ([235, Thm. 3.6]). *Let Assumptions 2.3.1–2.3.3 hold. Suppose that $\{\, u \in U_{ad} : \widehat{f}(u) \le \gamma \,\}$ is nonempty and bounded for some $\gamma \in \mathbb{R}$, and $U_{ad} \subset U$ is closed and convex, where $\widehat{f}$ is defined in* (2.1.9). *Then the DROP* (2.1.1) *has an optimal solution.*

We apply Lemma 2.3.7 to establish Theorem 2.3.6.

Lemma 2.3.7 ([235, Lem. 3.8]). *If Assumptions 2.3.1–2.3.3 hold, then $\widehat{f} : U_{ad} \to \mathbb{R}$ defined in* (2.1.9) *is weakly lower semicontinuous.*

Proof. First, we prove that $\widehat{f}$ is finite-valued. We have $\{\mathcal{N}(\mu,\Sigma) : \|\bar{\Sigma}^{-1/2}(\mu - \bar{\mu})\|_2 \le \Delta,\ \sigma_0\bar{\Sigma} \preccurlyeq \Sigma \preccurlyeq \sigma_1\bar{\Sigma}\,\} \subset \mathcal{P}$ [57, pp. 185–186] (see also [235, Lem. 3.7]) and, hence, $\mathcal{P} \ne \emptyset$. Fix $u \in U_{\mathrm{ad}}$ and $\delta > 0$. Since $\mathcal{P} \ne \emptyset$ and Assumption 2.3.3 holds, we obtain, for some $t > 0$ and all $P \in \mathcal{P}$, the estimate $\mathbb{E}_P[|\widehat{J}(u,\xi)|] \le t + \mathbb{E}_P[|\widehat{J}(u,\xi)|1_{|\widehat{J}(u,\xi)|\ge t}] \le t + \delta$. Hence $\widehat{f}(u) \in \mathbb{R}$. Since $u \in U_{\mathrm{ad}}$ is arbitrary, $\widehat{f}$ is finite-valued.
Now, fix $P \in \mathcal{P}$ and $(u_k) \subset U_{\mathrm{ad}}$ with $u_k \rightharpoonup u \in U_{\mathrm{ad}}$ as $k \to \infty$. We show that

$$\liminf_{k\to\infty} \mathbb{E}_P[\widehat{J}(u_k,\xi)] \ge \mathbb{E}_P[\widehat{J}(u,\xi)]. \tag{2.3.1}$$

Assumptions 2.3.1 and 2.3.2 imply that $\widehat{J}(\cdot,\xi)$ is weakly lower semicontinuous for all $\xi \in \mathbb{R}^p$ and $\widehat{J}(u,\cdot)$ is continuous for all $u \in U_{\mathrm{ad}}$. Hence $\widehat{J}(u,\cdot)$ is measurable for all $u \in U_{\mathrm{ad}}$ [169, Lem. 1.5]. Using [169, Lem. 1.9], we find that $\liminf_{k\to\infty} \widehat{J}(u_k,\cdot)$ is measurable. We deduce $\mathbb{E}_P[\liminf_{k\to\infty} \widehat{J}(u_k,\xi)] \geq \mathbb{E}_P[\widehat{J}(u,\xi)]$. Since $\widehat{J}(u,\xi) \geq \gamma$ for all $(u,\xi) \in U \times \mathbb{R}^p$, Fatou's lemma (see, e.g., [57, p. 232]) yields (2.3.1).
It must yet be shown that $\widehat{f}$ (see (2.1.9)) is weakly lower semicontinuous. Fix $\varepsilon > 0$. Since $\widehat{f}(u) < \infty$, there exists $P_\varepsilon \in \mathcal{P}$ with $\widehat{f}(u) \leq \mathbb{E}_{P_\varepsilon}[\widehat{J}(u,\xi)] + \varepsilon$. Now, (2.3.1) ensures

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

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

Proof of Theorem 2.3.6. Owing to Lemma 2.3.7, we can apply the direct method of the calculus of variations to prove that (2.1.1) has an optimal solution $u^* \in U_{\mathrm{ad}}$ (see also [46, Cor. 2.29]). □

We can establish the existence of optimal solution to the DROP (2.1.1) under slightly different hypotheses than those imposed by Assumptions 2.3.1–2.3.3.

Remark 2.3.8. Let Assumptions 2.3.1–2.3.3 hold, but instead of requiring the continuity of $S(u,\cdot) : \mathbb{R}^p \to Y$ for all $u \in U_{\mathrm{ad}}$, we require its strong measurability, and instead of imposing continuity of J, we impose the measurability of $J(\cdot,u,\cdot)$ for each $u \in U_{\mathrm{ad}}$. In this case, the following argumentation implies that, for each $u \in U_{\mathrm{ad}}$, the function $\widehat{J}(u,\cdot)$ defined in (2.1.3) is measurable. Since $S(u,\cdot)$ is strongly measurable, Pettis' measurability theorem [150, Thm. 3.5.3] implies that $\mathbb{R}^p \ni \xi \mapsto (S(u,\xi),\xi) \in Y \times \mathbb{R}^p$ is strongly measurable. Combined with the measurability of $J(\cdot,u,\cdot)$ and the composition rule [159, Cor. 1.1.11], we conclude that $\widehat{J}(u,\cdot) = J(\cdot,u,\cdot) \circ (S(u,\cdot),\cdot)$ is measurable.
Under these hypotheses, the proof of Lemma 2.3.7 may be modified to establish the weak lower semicontinuity of the function $\widehat{f} : U_{\mathrm{ad}} \to \mathbb{R}$ defined in (2.1.9). If Y is separable, then Assumptions 2.3.1–2.3.3 imply these modified conditions. Indeed, in this case, $S(u,\cdot) : \mathbb{R}^p \to Y$ is strongly measurable for all $u \in U_{\mathrm{ad}}$ (see [169, Lem. 1.5] and [150, Cor. 2 on p. 73]).

2.3.2 Existence of Worst-Case Distributions

We show that a worst-case distribution of the maximization problem in (2.1.1) exists. A worst-case distribution of (2.1.5) is the normal distribution, where the mean is a maximizer of (2.1.8) and the covariance matrix is one of (2.1.7).

Theorem 2.3.9 ([235, Thm. 3.9])**.** *If Assumptions 2.3.1–2.3.3 hold and $u \in U_{ad}$, then there exists a worst-case distribution of* (2.1.1)*.*

We use Lemmas 2.3.10 and 2.3.11 to prove Theorem 2.3.9. Lemmas 2.3.10 and 2.3.11 assert the weak-star sequential compactness of the ambiguity set $\mathcal{P}$ defined in (2.1.2). We say that the sequence $(P_k) \subset \mathcal{M}$ converges weakly to $P \in \mathcal{M}$ as $k \to \infty$, abbreviated with $P_k \Rightarrow P$ as $k \to \infty$, if, for each bounded, continuous function $f : \mathbb{R}^p \to \mathbb{R}$, we have $\mathbb{E}_{P_k}[f] \to \mathbb{E}_P[f]$ as $k \to \infty$ [169, p. 65], [45, Def. 1.4.1].

Lemma 2.3.10 ([235, Lem. 3.10])**.** *If $(P_k) \subset \mathcal{P}$ fulfills $P_k \Rightarrow P \in \mathcal{M}$ as $k \to \infty$, then $P \in \mathcal{P}$.*

Proof. Fix $i, j \in \{1,\dots,p\}$ and $d \in \mathbb{R}^p$. We define the continuous functions θ_1, θ_2, $\theta_3 : \mathbb{R}^p \to \mathbb{R}$ by $\theta_1(\xi) = \xi_i$, $\theta_2(\xi) = \xi_i\xi_j$, and $\theta_3(\xi) = \exp(d^T\xi)$. We show that these functions are uniformly integrable. We have $|\theta_1(\xi)| \leq \|\xi\|_2$ and $|\theta_2(\xi)| \leq \|\xi\|_2^2$ for all $\xi \in \mathbb{R}^p$. Lemma 2.9.1 and (2.1.2) imply $\sup_{P\in\mathcal{P}} \mathbb{E}_P[\|\xi\|_2^r] < \infty$ for $r = 1, 2, 4$. For all $P \in \mathcal{P}$, we have

$$\mathbb{E}_P[|\theta_3(\xi)|^2] = \mathrm{e}^{2d^T\mathbb{E}_P[\xi]}\mathbb{E}_P[\mathrm{e}^{2d^T(\xi-\mathbb{E}_P[\xi])}] \leq \mathrm{e}^{2d^T\mathbb{E}_P[\xi]+2\sigma_1 d^T\bar{\Sigma}d}. \tag{2.3.2}$$

Hence, Lemma 2.3.4 implies that θ_1, θ_2 and θ_3 are uniformly integrable. Combined with [45, Thm. 2.7.1], we find that $\mathbb{E}_{P_k}[\xi_i] = \mathbb{E}_{P_k}[\theta_1(\xi)] \to \mathbb{E}_P[\theta_1(\xi)] = \mathbb{E}_P[\xi_i]$, and $\mathbb{E}_{P_k}[\xi_i\xi_j] = \mathbb{E}_{P_k}[\theta_2(\xi)] \to \mathbb{E}_P[\theta_2(\xi)] = \mathbb{E}_P[\xi_i\xi_j]$ as $k \to \infty$. Since $i, j \in \{1, \ldots, p\}$ are arbitrary, we obtain

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

Combined with [45, Thm. 2.7.1], we get

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

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

Lemma 2.3.11 ([235, Lem. 3.11])**.** *If $(P_k) \subset \mathcal{P}$, then (P_k) has a weakly convergent subsequence $(P_k)_K$ such that $P_k \Rightarrow P \in \mathcal{M}$ as $K \ni k \to \infty$.*

Proof. We show that (P_k) is tight, that is, $\sup_{k\in\mathbb{N}_0} P_k(\|\xi\|_2 \geq r) \to 0$ as $r \to \infty$ [169, p. 85], [45, Def. 1.4.10]. Lemma 2.9.1 ensures $\sup_{k\in\mathbb{N}_0} \mathbb{E}_{P_k}[\|\xi\|_2^2] < \infty$. Markov's inequality gives $\sup_{k\in\mathbb{N}_0} P_k(\|\xi\|_2 > \sqrt{t}) \leq (\sup_{k\in\mathbb{N}_0} \mathbb{E}_{P_k}[\|\xi\|_2^2])/t \to 0$ as $t \to \infty$. Hence (P_k) is tight. Combined with [169, Lem. 5.20 and Prop. 5.21], we conclude that (P_k) has a subsequence $(P_k)_K$ with $P_k \Rightarrow P \in \mathcal{M}$ as $K \ni k \to \infty$. □

Proof of Theorem 2.3.9. Lemma 2.3.7 yields $\sup_{P\in\mathcal{P}} \mathbb{E}_P[\widehat{J}(u,\xi)] \in \mathbb{R}$. Let $(P_k) \subset \mathcal{P}$ satisfy $\lim_{k\to\infty} \mathbb{E}_{P_k}[\widehat{J}(u,\xi)] = \sup_{P\in\mathcal{P}} \mathbb{E}_P[\widehat{J}(u,\xi)]$. Lemma 2.3.11 implies that there exists a subsequence $(P_k)_K$ of (P_k) with $P_k \Rightarrow P^* \in \mathcal{M}$ as $K \ni k \to \infty$. Lemma 2.3.10 ensures $P^* \in \mathcal{P}$. Assumptions 2.3.1–2.3.3 imply that $\widehat{J}(u,\cdot)$ is continuous and uniformly integrable. Hence, the mapping theorem [45, Thm. 2.7.1] yields $\mathbb{E}_{P^*}[\widehat{J}(u,\xi)] = \lim_{K\ni k\to\infty} \mathbb{E}_{P_k}[\widehat{J}(u,\xi)] = \sup_{P\in\mathcal{P}} \mathbb{E}_P[\widehat{J}(u,\xi)]$. □

2.3.3 Existence of Optimal Solutions of the Approximated and Smoothed DROPs

We show that the approximated DROP (2.1.5) and the smoothed DROP (2.2.1) have optimal solutions under suitable assumptions.

Assumption 2.3.12 ([235, Assumption 3.12])**.** *For some $\epsilon > 0$ and each $(u,\xi) \in U_{ad} \times B_\epsilon(\bar{\mu})$, $S(u,\xi) \in Y$ is the unique solution to: Find $y \in Y$: $e(y,u,\xi) = 0$, where $S : U_{ad} \times B_\epsilon(\bar{\mu}) \to Y$ and $e : Y \times U \times B_\epsilon(\bar{\mu}) \to Z$.*

(a) *For all $u \in U_{ad}$, $J(\cdot,u,\cdot)$ is twice continuously differentiable, where $J : Y \times U_{ad} \times B_\epsilon(\bar{\mu}) \to \mathbb{R}$.*
(b) *The mapping $e : Y \times U \times B_\epsilon(\bar{\mu}) \to Z$ is twice continuously differentiable. For each $(u,\xi) \in U_{ad} \times B_\epsilon(\bar{\mu})$, the operator $e_y(S(u,\xi),u,\xi) \in \mathscr{L}(Y,Z)$ is boundedly invertible.*
(c) *The function $\widehat{J}(\cdot,\bar{\mu}) : U_{ad} \to \mathbb{R}$ is weakly lower semicontinuous, and $\nabla_\xi\widehat{J}(\cdot,\bar{\mu}) : U_{ad} \to \mathbb{R}^p$ and $\nabla_{\xi\xi}\widehat{J}(\cdot,\bar{\mu}) : U_{ad} \to \mathbb{S}^p$ are weakly(-strongly) continuous.*

Assumptions 2.3.12 (a) and 2.3.12 (b) imply that the objective functions of the approximated DROP (2.1.5) and the smoothed DROP (2.2.1) are well-defined. Assumption 2.3.12 (c) may be verified using Lemma 2.3.15 or Lemma 2.3.16. Assumption 2.3.12 implies that the approximated DROP (2.1.5) and the smoothed DROP (2.2.1) have optimal solutions.

Theorem 2.3.13 ([235, Thm. 3.13])**.** *Let Assumption 2.3.12 hold, and let $U_{ad} \subset U$ be nonempty, closed and convex. Suppose that U_{ad} is bounded or the function F defined in* (2.1.6) *is coercive. Then the approximated DROP* (2.1.5) *has an optimal solution and, for each $(\tau_k,\nu_k,\eta_k) \in \mathbb{R}^3_{++}$, the smoothed DROP* (2.2.1) *has an optimal solution.*

We prove Theorem 2.3.13 using Lemma 2.3.14.

Lemma 2.3.14 ([235, Lem. 3.14])**.** *Let Assumption 2.3.12 hold and fix $t^k = (\tau_k, \nu_k, \eta_k) \in \mathbb{R}^3_{++}$. Then $F : U_{ad} \to \mathbb{R}$ and $\widetilde{F}(\cdot; t^k) : U_{ad} \to \mathbb{R}$ are weakly lower semicontinuous.*

Proof. The mapping λ is (Lipschitz) continuous [157, Cor. 6.3.8]. Combined with Assumption 2.3.12 (c) and the definition of G (see (2.2.4)) and $E(\cdot;\eta_k)$ (see (1.4.10)), we find that $\lambda \circ G : U_{\mathrm{ad}} \to \mathbb{R}$ and $E(\cdot;\eta) \circ G : U_{\mathrm{ad}} \to \mathbb{R}$ are weakly continuous. From (2.2.3) and (2.2.5), we obtain that $\varphi : U_{\mathrm{ad}} \to \mathbb{R}$ and $\widetilde{\varphi}(\cdot;\tau_k) : U_{\mathrm{ad}} \to \mathbb{R}$ are weakly continuous. Owing to Assumption 2.3.12, [148, Thm. 2.5] implies that $\psi : U_{\mathrm{ad}} \to \mathbb{R}$ (see (2.1.8)) and $\widetilde{\psi}(\cdot;\nu_k,\eta_k) : U_{\mathrm{ad}} \to \mathbb{R}$ (see (2.2.7)) are weakly lower semicontinuous. The weak lower semicontinuity of $\widehat{J}(\cdot,\bar{\mu}) : U_{\mathrm{ad}} \to \mathbb{R}$ implies that of $F : U_{\mathrm{ad}} \to \mathbb{R}$ and of $\widetilde{F}(\cdot;t^k) : U_{\mathrm{ad}} \to \mathbb{R}$. □

Proof of Theorem 2.3.13. Lemma 2.3.14 yields the lower semicontinuity of F and $\widetilde{F}(\cdot;t^k)$. If F is coercive, then (2.2.6) and (2.2.10) imply that $\widetilde{F}(\cdot;t^k)$ is coercive. Now, the direct method of the calculus of variations yields the existence of an optimal solution of (2.1.5) and of (2.2.1). □

Assumption 2.3.12 (c) may be verified using compact embeddings.

Lemma 2.3.15 ([235, Rem. 3.15])**.** *Let $\widetilde{U}$ be a Banach space and let $U \hookrightarrow \widetilde{U}$ be a compact embedding, and let $J_2 : U \to \mathbb{R}$ be weakly lower semicontinuous. Consider $J_1 : Y \times B_\epsilon(\bar{\mu}) \to \mathbb{R}$ and the solution operator $S : \widetilde{U} \times B_\epsilon(\bar{\mu}) \to Y$, where $\varepsilon > 0$. Let J_1 and S be twice continuously differentiable. Suppose that $J(y,u,\xi) = J_1(y,\xi) + J_2(u)$ for all $(y,u,\xi) \in Y \times U \times B_\epsilon(\bar{\mu})$. Then Assumptions 2.3.12* (a) *and 2.3.12* (c) *hold.*

Proof. By assumption, the function $\widehat{J}(\cdot,\bar{\mu}) : U_{\mathrm{ad}} \to \mathbb{R}$ defined in (2.1.3) is weakly lower semicontinuous. Moreover, $\widetilde{U} \times B_\epsilon(\bar{\mu}) \ni (u,\xi) \mapsto J_1(S(u,\xi),\xi)$ is twice continuously differentiable, and we have $\mathrm{D}_\xi \widehat{J} = \mathrm{D}_\xi(J(S(\cdot,\cdot),\cdot))$ and $\mathrm{D}_{\xi\xi}\widehat{J} = \mathrm{D}_{\xi\xi}(J(S(\cdot,\cdot),\cdot))$. We deduce the weak continuity of $\nabla_\xi \widehat{J}(\cdot,\bar{\mu}) : U_{\mathrm{ad}} \to \mathbb{R}^p$ and $\nabla_{\xi\xi}\widehat{J}(\cdot,\bar{\mu}) : U_{\mathrm{ad}} \to \mathbb{S}^p$. □

We verify Assumption 2.3.12 (c) for the DRO the unsteady Burgers' equation using Lemma 2.3.16 in section 2.7.2.

Lemma 2.3.16. *Let Assumptions 2.3.12* (a) *and 2.3.12* (b) *hold. Consider the setting of Lemma 2.3.5 and let $Y \hookrightarrow H$ be a compact embedding. Suppose that Assumption 2.3.1* (a) *holds, and $S_\xi(\cdot,\bar{\mu})s_\xi : U_{ad} \to Y$ and $S_{\xi\xi}(\cdot,\bar{\mu})[s_\xi,s_\xi] : U_{ad} \to Y$ are weakly-weakly continuous for all $s_\xi \in \mathbb{R}^p$. Then Assumption 2.3.12* (c) *holds.*

Proof. The tracking-type function J is convex and continuous and, hence, it is weakly lower semicontinuous [151, Thm. 1.18]. Combined with Assumption 2.3.1 (a), we find that $\widehat{J} : U_{\mathrm{ad}} \to \mathbb{R}$ is weakly lower semicontinuous.
We have, for all y, $s_y \in Y$ and $u \in U$,

$$\langle J_y(y,u,\bar{\mu}), s_y\rangle_{Y^*,Y} = (y - y_d, s_y)_H, \quad \text{and} \quad \langle J_{yy}(y,u,\bar{\mu})s_y, s_y\rangle_{Y^*,Y} = (s_y,s_y)_H. \tag{2.3.3}$$

Using J_ξ, $J_{\xi\xi}$, $J_{\xi y} = 0$, we get, for all $u \in U$ and $s_\xi \in \mathbb{R}^p$,

$$\begin{aligned} \langle \widehat{J}_\xi(u,\bar{\mu}), s_\xi\rangle_{(\mathbb{R}^p)^*,\mathbb{R}^p} &= \langle J_y(S(u,\bar{\mu}),u,\bar{\mu}), S_\xi(u,\bar{\mu})s_\xi\rangle_{Y^*,Y}, \\ \langle \widehat{J}_{\xi\xi}(u,\bar{\mu})s_\xi, s_\xi\rangle_{(\mathbb{R}^p)^*,\mathbb{R}^p} &= \langle J_{yy}(S(u,\bar{\mu}),u,\bar{\mu})S_\xi(u,\bar{\mu})s_\xi, S_\xi(u,\bar{\mu})s_\xi\rangle_{Y^*,Y} \\ &\quad + \langle J_y(S(u,\bar{\mu}),u,\bar{\mu}), S_{\xi\xi}(u,\bar{\mu})[s_\xi,s_\xi]\rangle_{Y^*,Y}. \end{aligned} \tag{2.3.4}$$

Now, fix $s_\xi \in \mathbb{R}^p$, and fix $(u_k) \subset U_{\mathrm{ad}}$ with $u_k \rightharpoonup u \in U_{\mathrm{ad}}$ as $k \to \infty$.

The compact embedding $Y \hookrightarrow H$, the boundedness of $(\|S(u_k,\bar\mu)\|_Y)$ [46, Thm. 2.23], the weak-weak continuity of $S(\cdot,\bar\mu)$ and of $S_\xi(\cdot,\bar\mu)s_\xi$, and the Cauchy–Schwarz inequality yield $|(S(u_k,\bar\mu), S_\xi(u_k,\bar\mu)s_\xi - S_\xi(u,\bar\mu)s_\xi)_H| \le \|S(u_k,\bar\mu)\|_H \|S_\xi(u_k,\bar\mu)s_\xi - S_\xi(u,\bar\mu)s_\xi\|_H \to 0$ as $k\to\infty$ [151, Lem. 1.6]. Combined with (2.3.3) and (2.3.4), we deduce

$$\begin{aligned}\langle \widehat{J}_\xi(u_k,\bar\mu), s_\xi\rangle_{(\mathbb{R}^p)^*,\mathbb{R}^p} &= (S(u_k,\bar\mu)-y_d, S_\xi(u,\bar\mu)s_\xi)_H + (S(u_k,\bar\mu)-y_d, S_\xi(u_k,\bar\mu)s_\xi - S_\xi(u,\bar\mu)s_\xi)_H\\ &\to \langle \widehat{J}_\xi(u,\bar\mu), s_\xi\rangle_{(\mathbb{R}^p)^*,\mathbb{R}^p}.\end{aligned}$$

Hence $\nabla_\xi \widehat{J}(\cdot,\bar\mu): U_{\mathrm{ad}} \to \mathbb{R}^p$ is weakly(-strongly) continuous.
It must yet be shown that $\nabla_{\xi\xi}\widehat{J}(\cdot,\bar\mu): U_{\mathrm{ad}} \to \mathbb{S}^p$ is weakly(-strongly) continuous. Since $Y \hookrightarrow H$ is compact and $S_\xi(\cdot,\bar\mu)s_\xi : U_{\mathrm{ad}} \to Y$ is weakly-weakly continuous, we have $\|S_\xi(u_k,\bar\mu)s_\xi\|_H^2 \to \|S_\xi(u,\bar\mu)s_\xi\|_H^2$ as $k\to\infty$. Using a similar reasoning as above and the weak-weak continuity of $S_{\xi\xi}(\cdot,\bar\mu)[s_\xi,s_\xi]$, we find that

$$\begin{aligned}\langle J_y(S(u_k,\bar\mu),u,\bar\mu), S_{\xi\xi}(u_k,\bar\mu)[s_\xi,s_\xi]\rangle_{Y^*,Y} &= (S(u_k,\bar\mu)-y_d, S_{\xi\xi}(u_k,\bar\mu)[s_\xi,s_\xi])_H\\ &\to \langle J_y(S(u,\bar\mu),u,\bar\mu), S_{\xi\xi}(u,\bar\mu)[s_\xi,s_\xi]\rangle_{Y^*,Y}.\end{aligned}$$

Putting together the pieces, we conclude that, for all $s_\xi \in \mathbb{R}^p$, $\langle \widehat{J}_{\xi\xi}(u_k,\bar\mu)s_\xi, s_\xi\rangle_{(\mathbb{R}^p)^*,\mathbb{R}^p} \to \langle \widehat{J}_{\xi\xi}(u,\bar\mu)s_\xi, s_\xi\rangle_{(\mathbb{R}^p)^*,\mathbb{R}^p}$ as $k\to\infty$. Combined with $(\nabla_{\xi\xi}\widehat{J}(u_k,\bar\mu)) \subset \mathbb{S}^p$ and $\nabla_{\xi\xi}\widehat{J}(u,\bar\mu)\in\mathbb{S}^p$, we find that, for each $i,\, j \in \{1,\dots,p\}$,

$$\begin{aligned}2e_i^T\nabla_{\xi\xi}\widehat{J}(u_k,\bar\mu)e_j &= \langle \widehat{J}_{\xi\xi}(u_k,\bar\mu)[e_i+e_j], e_i+e_j\rangle_{(\mathbb{R}^p)^*,\mathbb{R}^p}\\ &\quad - \langle \widehat{J}_{\xi\xi}(u_k,\bar\mu)e_j, e_j\rangle_{(\mathbb{R}^p)^*,\mathbb{R}^p} - \langle \widehat{J}_{\xi\xi}(u_k,\bar\mu)e_i, e_i\rangle_{(\mathbb{R}^p)^*,\mathbb{R}^p} \to 2e_i^T\nabla_{\xi\xi}\widehat{J}(u,\bar\mu)e_j.\end{aligned}$$

Here, e_i is the ith canonical unit vector of $\mathbb{R}^p$. We deduce the weak continuity of $\nabla_{\xi\xi}\widehat{J}(\cdot,\bar\mu)$. □

2.4 Convergence of the Smoothing Method

We present a convergence result for a sequence of optimal solutions of the smoothed DROPs (2.2.1) as the smoothing parameters converge to zero. Theorem 2.4.1 implies the global convergence of a sequence of minimizers generated by Algorithm 2 to an optimal solution of the approximated DROP (2.1.5).

Theorem 2.4.1 ([235, Thm. 4.1])**.** *Let the conditions of Theorem 2.3.13 hold, and let $t^k = (\tau_k,\nu_k,\eta_k) \in \mathbb{R}^3_{++}$ fulfill $t^k \to 0$ as $k\to\infty$. Suppose that, for each $k\in\mathbb{N}_0$, u_k is an optimal solution of (2.2.1). Then (u_k) is bounded, and each weak limit point of (u_k) is an optimal solution of (2.1.5).*

Proof. We fix $u\in U_{\mathrm{ad}}$ and $k\in\mathbb{N}_0$. Using (2.2.6), (2.2.10), and $\widetilde{F}(u_k;t^k) \le \widetilde{F}(u;t^k)$, we obtain

$$F(u_k) \le \widetilde{F}(u_k;t^k) \le \widetilde{F}(u;t^k) \le F(u) + \tfrac{1}{2}\tau_k p\ln 2 + 2\sqrt{2\nu_k}\Delta + \tfrac{1}{2}\Delta^2\eta_k\ln p, \tag{2.4.1}$$

where F is defined in (2.1.6) and $\widetilde{F}$ in (2.2.1). Since either U_{ad} is bounded or F is coercive and (2.4.1) holds, the sequence $(u_k)\subset U_{\mathrm{ad}}$ is bounded.
Let u^* be a weak accumulation point of (u_k). Then there exists $(u_k)_K \subset (u_k)$ such that $u_k \rightharpoonup u^*$ as $K \ni k\to\infty$. Since U_{ad} is closed and convex, we have $u^*\in U_{\mathrm{ad}}$ [46, Thm. 2.23]. Lemma 2.3.14 yields $F(u^*) \le \liminf_{K\ni k\to\infty} F(u_k)$. Combined with (2.4.1) and $t^k\to 0$ as $k\to\infty$, we obtain $F(u^*)\le F(u)$. Consequently, u^* is an optimal solution of (2.1.5). □

2.5 Error of Quadratic Approximation

We show that the worst-case expected error between the objective function of (2.1.1) and that of (2.1.5) converges to zero for "shrinking" ambiguity sets.

Lemma 2.5.1 ([235, Lem. 5.1])**.** *Let Assumptions 2.3.12* (a) *and 2.3.12* (b) *hold, and let* $u \in U_{ad}$*. Suppose that* $L(u,\cdot) : \mathbb{R}^p \to \mathbb{R}_+$ *is measurable such that* $\sup_{P\in\mathcal{P}} \mathbb{E}_P[L(u,\xi)^2] < \infty$*, and*

$$|\widehat{J}(u,\xi) - Q(u,\xi;\bar{\mu})| \le (L(u,\xi)/6)\|\xi - \bar{\mu}\|_2^3, \quad \text{for all} \quad \xi \in \mathbb{R}^p, \tag{2.5.1}$$

where $\widehat{J}$ *is defined in* (2.1.3) *and* Q *in* (2.1.4)*. Then*

$$\sup_{P\in\mathcal{P}} \mathbb{E}_P[|\widehat{J}(u,\xi) - Q(u,\xi;\bar{\mu})|] \to 0 \quad \text{as} \quad (\Delta,\sigma_1) \to 0^+. \tag{2.5.2}$$

Proof. Fix $P \in \mathcal{P}$. Using Hölder's inequality and (2.5.1), we find that

$$\mathbb{E}_P[|\widehat{J}(u,\xi) - Q(u,\xi;\bar{\mu})|] \le (1/6)(\mathbb{E}_P[|L(u,\xi)|^2])^{1/2}(\mathbb{E}_P[\|\xi - \bar{\mu}\|_2^6])^{1/2}. \tag{2.5.3}$$

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

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

Lemma 2.9.1 yields $\mathbb{E}_P[\|\xi - \mathbb{E}_P[\xi]\|_2^6] \le 2(6/\mathrm{e})^6(I \bullet \sigma_1\bar{\Sigma})^6$. Using (2.1.2), we obtain $\|\bar{\Sigma}^{-1/2}(\mathbb{E}_P[\xi] - \bar{\mu})\|_2 \le \Delta$ and, hence, $\|\mathbb{E}_P[\xi] - \bar{\mu}\|_2^6 \le \|\bar{\Sigma}^{1/2}\|_2^6\Delta^6$. We deduce

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

Hence $\sup_{P\in\mathcal{P}} \mathbb{E}_P[\|\xi - \bar{\mu}\|_2^6] \to 0$ as $(\Delta,\sigma_1) \to 0^+$. Combined with (2.5.3), we obtain (2.5.2). □

2.6 Evaluation of Smoothing Functions and their Derivatives

We derive formulas for the derivative of the smoothing functions $\widetilde{\varphi}$ (see (2.2.5)) and $\widetilde{\psi}$ (see (2.2.7)). Throughout the section, let Assumptions 2.3.12 (a) and 2.3.12 (b) be satisfied, and fix $\bar{u} \in U$. Moreover, let $J(y,\cdot,\xi)$ be continuously differentiable in a neighborhood of $\bar{u}$ for all $(y,\xi) \in Y \times \mathbb{R}^p$, and fix $(\tau,\nu,\eta) \in \mathbb{R}^3_{++}$.

2.6.1 Smoothing Function of the SDP

In order to evaluate $\widetilde{\varphi}$, we propose to compute the Hessian matrix $\nabla_{\xi\xi}J(\bar{u},\bar{\mu})$ when the number of parameters p is moderate. We use the identity $G(\bar{u}) \bullet I = \nabla_{\xi\xi}J(\bar{u},\bar{\mu}) \bullet \bar{\Sigma}$ to evaluate the first addend in (2.2.5) and compute an eigendecomposition of $G(\bar{u})$ via the (generalized) eigenvalue problem $\nabla_{\xi\xi}J(u,\bar{\mu})q = \lambda\bar{\Sigma}^{-1}q$ with $q \neq 0$, where G is defined in (2.2.4). For each $s \in U$ and $P \in \mathbb{S}^p$, we have

$$\langle \mathrm{D}G(\bar{u})^*P, s\rangle_{U^*,U} = P \bullet (\mathrm{D}G(\bar{u})s) = \langle \mathrm{D}(G(\bar{u}) \bullet P), s\rangle_{U^*,U}. \tag{2.6.1}$$

From section 2.2.1, [217, Lem. 3.1], (2.2.5) and (2.6.1), we obtain that

$$\mathrm{D}_u\widetilde{\varphi}(\bar{u};\tau) = (\sigma_0/2)\mathrm{D}G(\bar{u})^*I + ((\sigma_1 - \sigma_0)/2)\mathrm{D}G(\bar{u})^*[Q(\bar{u})M(\bar{u})Q(\bar{u})^T], \tag{2.6.2}$$

where $\widetilde{w}$ is defined in (1.3.3) and G in (2.2.4), $Q(\bar{u}) \in \mathbb{R}^{p\times p}$ fulfills $Q(\bar{u})^TQ(\bar{u}) = I$, and

$$G(\bar{u}) = Q(\bar{u})\mathrm{Diag}(\lambda(G(\bar{u})))Q(\bar{u})^T \quad \text{and} \quad M(\bar{u}) = \mathrm{Diag}(\nabla_x\widetilde{w}(\lambda(G(\bar{u}));\tau)). \tag{2.6.3}$$

Using (2.6.1), the matrix $\mathrm{D}G(\bar{u})^*[Q(\bar{u})M(\bar{u})Q(\bar{u})^T]$ in (2.6.2) becomes

$$\mathrm{D}G(\bar{u})^*[Q(\bar{u})M(\bar{u})Q(\bar{u})^T] = \sum_{i=1}^p m_{ii}(\bar{u})\,\mathrm{D}_u(q_i(\bar{u})^TG(u)q_i(\bar{u}))\big|_{u=\bar{u}}\,, \tag{2.6.4}$$

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

2.6.2 Smoothing Function of the TRP

In order to evaluate $\widetilde{\psi}$, we compute $E(G(\bar{u});\eta)$ (see (1.4.10)) using the eigenvalues of $G(\bar{u})$ that are used to evaluate $\widetilde{\varphi}$; see section 2.6.1. From section 2.2.2 and [46, Rem. 4.14], we obtain that

$$\mathrm{D}_u\widetilde{\psi}(\bar{u};\nu,\eta) = \mathrm{D}_u(g(\bar{u})^T s^*) + \tfrac{1}{2}\mathrm{D}_u((s^*)^T G(\bar{u}) s^*) + \tfrac{1}{2}\widetilde{s}_{p+2}^2 \mathrm{D}_u E(G(\bar{u});\eta), \tag{2.6.5}$$

where $\widetilde{s} = (s^*, \tilde{s}_{p+1}, \tilde{s}_{p+2}) \in \mathbb{R}^{p+2}$ is the optimal solution of (2.2.7) for $u = \bar{u}$. We have

$$\nabla_A E(A;\eta) = R(A)\mathrm{Diag}(\theta(A))R(A)^T \quad \text{and} \quad \theta_i(A;\eta) = \frac{\exp(\lambda_i(A)/\eta)}{\sum_{i=1}^p \exp(\lambda_i(A)/\eta)}, \tag{2.6.6}$$

where $R(A) \in \mathbb{R}^{p\times p}$, $R(A)^T R(A) = I$, and $A = R(A)\mathrm{Diag}(\lambda(A))R(A)^T \in \mathbb{S}^p$ [247, eq. (18)] (see also section 1.4.3). Combined with the chain rule and (2.6.1), we find that

$$\mathrm{D}_u E(G(\bar{u});\eta) = \sum_{i=1}^p \theta_i(\bar{u})\,\mathrm{D}_u(q_i(\bar{u})^T G(u) q_i(\bar{u}))\big|_{u=\bar{u}}, \tag{2.6.7}$$

where $\theta_i(\bar{u}) = \theta_i(G(\bar{u});\eta)$ and $q_i(\bar{u})$ is the ith column of the matrix $Q(\bar{u})$ defined in (2.6.3).

2.7 Applications and Numerical Results

We formulate and analyze two DROPs with nonlinear PDEs, and present numerical results.

2.7.1 DRO of Steady Burgers' Equation

We formulate an optimal control problem of a parameterized steady Burgers' equation that was studied in [191, 188, 185] for risk-averse objective functions other than (2.1.9). We consider

$$\min_{u\in U} \sup_{P\in\mathcal{P}} \mathbb{E}_P[(1/2)\|S(u,\xi) - y_d\|_{L^2(\mathcal{D})}^2] + (\alpha/2)\|u\|_{L^2(\mathcal{D})}^2, \tag{2.7.1}$$

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

$$\begin{aligned} -\kappa(\xi)y_{xx}(x) + y(x)y_x(x) &= \xi_2/100 + u(x), \qquad x \in \mathcal{D},\\ y(0) = 1 + \xi_3/1000, \quad y(1) &= \xi_4/1000, \end{aligned} \tag{2.7.2}$$

where $p = 4$, $\xi \in \mathbb{R}^p$, $u \in U$ and $\kappa : \mathbb{R}^p \to \mathbb{R}_{++}$, $\kappa(\xi) = 10^{\xi_1 - 2}$. We refer the reader to [328, 329] for the analysis of deterministic control problems subject to the steady Burgers' equation.
We define $V = H_0^1(\mathcal{D})$ and $e = (e_1, e_2) : H^1(\mathcal{D}) \times V^* \times \mathbb{R}^p \to V^* \times \mathbb{R}^2$ by[1]

$$\langle e_1(y,u,\xi), v\rangle_{V^*,V} = \int_{\mathcal{D}} [\kappa(\xi)y_x(x)v_x(x) + (y(x)y_x(x) - \tfrac{\xi_2}{100})v(x)]\mathrm{d}x - \langle u, v\rangle_{V^*,V} \tag{2.7.3}$$

for all $v \in V$ and $e_2(y,u,\xi) = (y(0) - 1 - \xi_3/1000, y(1) - \xi_4/1000)$; cf. [328, pp. 71 and 79].
Our computational results presented below show that our approximation scheme produces controls with similar behavior as those obtained in [191, sect. 6.2] via the minimization of the superquantile/conditional value-at-risk using the sample average approximation, while our scheme requires fewer PDE solutions.
We show that the differentiability requirements in Assumption 2.3.12 are fulfilled.

[1] By definition, we have $H_0^1(\mathcal{D})^* = H^{-1}(\mathcal{D})$ [151, p. 23]. We identify $L^2(\mathcal{D})$ with its dual. The embedding $U = L^2(\mathcal{D}) \hookrightarrow V^* = H^{-1}(\mathcal{D})$ is compact (see, e.g., [1, Thm. 6.2] and [196, Thm. 8.2-5]) and is given by $\langle v, w\rangle_{V^*,V} = (v,w)_{L^2(\mathcal{D})}$ for all $v \in U$ and $w \in V$ [151, pp. 39–40].

Lemma 2.7.1. *The operator e defined in* (2.7.3) *is twice continuously differentiable.*

Proof. For each fixed $\xi \in \mathbb{R}^p$, the derivations in [328, p. 81] imply that $e(\cdot,\cdot,\xi)$ is twice continuously differentiable. In light of the calculus rules [97, Thms. 8.9.1 and 8.12.6], it suffices to show that $e_1(y,u,\cdot)$ and $e_2(y,u,\cdot)$ are twice continuously differentiable for each $(y,u) \in Y \times V^*$. We fix $(y,u) \in Y \times V^*$. The mapping $e_2(y,u,\cdot) : \mathbb{R}^2 \to \mathbb{R}^2$ is affine and, hence, it is infinitely many times continuously differentiable. Next, we show that, for each $\xi \in \mathbb{R}^p$,

$$\langle \mathrm{D}_\xi e_1(y,u,\xi)h, v\rangle_{V^*,V} = \int_{\mathcal{D}} [\ln(10)10^{\xi_1-2}h_1 y'v' - (h_2/100)v]\mathrm{d}x \quad \text{for all} \quad v \in V,\, h \in \mathbb{R}^p.$$

For each $v \in V = H_0^1(\mathcal{D})$ with $\|v\|_V \leq 1$ and all $h \in \mathbb{R}^p$, the Cauchy–Schwarz inequality and $\|w'\|_{L^2(\mathcal{D})} \leq \|w\|_V$, valid for all $w \in V$, ensure

$$|\langle e_1(y,u,\xi+h) - e_1(y,u,\xi) - \mathrm{D}_\xi e_1(y,u,\xi)h, v\rangle_{V^*,V}| \leq |\kappa(\xi+h) - \kappa(\xi) - \mathrm{D}\kappa(\xi)h|\|y\|_Y.$$

Hence $e_1(y,u,\cdot)$ is Fréchet differentiable. A similar derivation, when combined with

$$\langle \mathrm{D}_{\xi\xi} e_1(y,u,\xi)[h,s], v\rangle_{V^*,V} = \int_{\mathcal{D}} [\ln(10)^2 10^{\xi_1-2}h_1 s_1 y'v']\mathrm{d}x \quad \text{for all} \quad v \in V,\, h,s \in \mathbb{R}^p,$$

implies that $e_1(y,u,\cdot)$ is twice Fréchet differentiable. The above formula also reveals the continuity of $\mathrm{D}_{\xi\xi}e_1(y,u,\cdot) \in \mathscr{L}(\mathbb{R}^p, \mathscr{L}(\mathbb{R}^p, V^*))$. □

The function $J : Y \times U \times \mathbb{R}^p \to \mathbb{R}_+$ defined by $J(y,u,\xi) = (1/2)\|y - y_d\|^2_{L^2(\mathcal{D})} + (\alpha/2)\|u\|^2_{L^2(\mathcal{D})}$ is convex and infinitely many times continuously differentiable. Hence Assumption 2.3.2 holds. Combined with Lemma 2.7.1, we find that the differentiability requirements in Assumption 2.3.12 are met.

We show that the parameterized steady Burgers' equation (2.7.2) has a solution based on the argumentation used in [188, sect. 5.2.1] and in [328, Chap. 4]. We fix $\xi \in \mathbb{R}^p$ and define $\varepsilon : \mathbb{R}^p \to \mathbb{R}$ by $\varepsilon(\xi) = \kappa(\xi)/2$.[2] From [328, Lem. 2.2 (p. 71)], we deduce the existence of $y_0(\xi) \in H^1(\mathcal{D})$ with $e_2(y_0(\xi),u,\xi) = 0$ for all $u \in V^*$, and $\|y_0(\xi)\|_{L^2(\mathcal{D})} \leq \varepsilon(\xi)$. The derivations in [328, p. 76] imply that there exists a solution $S(u,\xi) \in Y$ of the weak form of (2.7.2) for each $u \in V^*$. If $\kappa(\xi)$ is sufficiently large, then this solution is unique [328, Thm. 2.13 (p. 76)].

We prove that, for fixed $u \in V^*$, the set-valued solution mapping $\mathcal{S}(u,\cdot) : \mathbb{R}^p \rightrightarrows Y$ defined by $\mathcal{S}(u,\xi) = \{\, y \in Y : e(y,u,\xi) = 0 \,\}$ is measurable and there exists a measurable selection $S(u,\cdot)$ of $\mathcal{S}(u,\cdot)$ with $e(S(u,\xi),u,\xi) = 0$ for all $\xi \in \mathbb{R}^p$. In order to establish the assertions, we verify the hypotheses of the theorem on the measurablity of implicit multifunctions [66, Thm. III.38]. We recall that the separable Banach spaces Y, $\mathbb{R}^p$ and $V^* \times \mathbb{R}^2$ are equipped with their Borel-σ-field.[3] The continuity of e (see Lemma 2.7.1) implies that $e(\cdot,u,\cdot)$ is $\mathcal{B}(Y \times \mathbb{R}^p)$-$\mathcal{B}(V^* \times \mathbb{R}^2)$-measurable [169, Lem. 1.5]. Moreover, we have $\mathcal{B}(Y \times \mathbb{R}^p) = \mathcal{B}(Y) \otimes \mathcal{B}(\mathbb{R}^p)$ [169, Lem. 1.2], and the parameterized steady Burgers' equation (2.7.2) has a solution. Hence $\mathcal{S}(u,\xi)$ is nonempty for all $\xi \in \mathbb{R}^p$. Combining these statements with [66, Thm. III.38] yields the assertions (see also [66, pp. 80 and 86]).

Lemma 2.7.2. *If $u \in V^*$ and $S(u,\cdot) : \mathbb{R}^p \to Y$ is a measurable selection of $\mathcal{S}(u,\cdot)$, then $\|S(u,\cdot)\|^2_{L^2(\mathcal{D})}$ is uniformly integrable.*

[2] We choose $n = n(\xi) = ((1+10^{-3}\xi_3)^2 + 10^{-6}\xi_4^2)/(4\varepsilon(\xi)) + 1 \geq 1$ in the proof of [328, Lem. 2.2 (p. 71)].

[3] The Hilbert space $V = H_0^1(\mathcal{D})$ is separable and, hence, its dual $V^* = H^{-1}(\mathcal{D})$ is separable [196, pp. 242–243] (see also [1, Thm. 1.14]).

Proof. Since $Y = H^1(\mathcal{D}) \subset L^2(\mathcal{D})$, and $H^1(\mathcal{D})$ and $L^2(\mathcal{D})$ are separable, the measurability of $S(u,\cdot)$ implies that of $\|S(u,\cdot)\|_{L^2(\mathcal{D})}$ and of $\|S(u,\cdot)\|_Y$ [150, Thm. 3.5.2].
Owing to Lemma 2.3.4 and $\|S(u,\cdot)\|_{L^2(\mathcal{D})} \leq \|S(u,\cdot)\|_Y$, it suffices to show that $\|S(u,\cdot)\|_Y^2$ is uniformly integrable. To establish the uniform integrability of $\|S(u,\cdot)\|_Y^2$, we use a stability estimate for the solution to the Burgers' equation derived by Volkwein [328]. We fix $\xi \in \mathbb{R}^p$ and define $\widetilde{f} = \widetilde{f}(\xi) \in H^{-1}(\mathcal{D})$ by

$$\langle \widetilde{f}, \varphi \rangle_{H_0^1(\mathcal{D})^*, H_0^1(\mathcal{D})} = \langle u, \varphi \rangle_{H_0^1(\mathcal{D})^*, H_0^1(\mathcal{D})} - \int_{\mathcal{D}} [\kappa(\xi) y_0(\xi)' \varphi' + y_0(\xi) y_0(\xi)' \varphi] \mathrm{d}x \quad \text{for all} \quad \varphi \in H_0^1(\mathcal{D}),$$

cf. [328, p. 71]. We derive an upper bound on $\|\widetilde{f}\|_{H^{-1}(\mathcal{D})}$. Using the definition of the operator norm $\|\cdot\|_{H^{-1}(\mathcal{D})}$ and the Hölder inequality, we find that, for all $\varphi \in H_0^1(\mathcal{D})$,

$$\begin{aligned} |\langle \widetilde{f}, \varphi \rangle_{H_0^1(\mathcal{D})^*, H_0^1(\mathcal{D})}| \leq{}& \|u\|_{H^{-1}(\mathcal{D})} \|\varphi\|_{H_0^1(\mathcal{D})} + \kappa(\xi) \|y_0(\xi)'\|_{L^2(\mathcal{D})} \|\varphi'\|_{L^2(\mathcal{D})} \\ &+ \|y_0(\xi)'\|_{L^2(\mathcal{D})} \|y_0(\xi)\|_{L^2(\mathcal{D})} \|\varphi\|_{L^\infty(\mathcal{D})}. \end{aligned}$$

Combined with the definition of the $H^{-1}(\mathcal{D})$-norm and that of the $H^1(\mathcal{D})$-norm, and using $\|\varphi\|_{L^\infty(\mathcal{D})} \leq \|\varphi\|_{H_0^1(\mathcal{D})}$, valid for all $\varphi \in H_0^1(\mathcal{D})$ (see, e.g., [328, Lem. 3.4 (p. 9)]), we obtain

$$\|\widetilde{f}\|_{H^{-1}(\mathcal{D})} \leq \|u\|_{H^{-1}(\mathcal{D})} + \kappa(\xi) \|y_0(\xi)\|_Y + \|y_0(\xi)\|_Y \|y_0(\xi)\|_{L^2(\mathcal{D})} \tag{2.7.4}$$

Using [328, Lem. 2.3 (p. 72)], we get, with $\varepsilon(\xi) = \kappa(\xi)/2$ and $\kappa(\xi) = 10^{\xi_1 - 2}$,

$$\|S(u,\xi)\|_Y \leq (\sqrt{8}/\kappa(\xi)) \|\widetilde{f}\|_{H^{-1}(\mathcal{D})} + \|y_0(\xi)\|_Y.$$

Combined with (2.7.4) and $\sqrt{8} \leq 3$, we find

$$\|S(u,\xi)\|_Y \leq 4\|y_0(\xi)\|_Y + (3/\kappa(\xi)) \|u\|_{H^{-1}(\mathcal{D})} + (3/\kappa(\xi)) \|y_0(\xi)\|_Y \|y_0(\xi)\|_{L^2(\mathcal{D})}. \tag{2.7.5}$$

Using the derivations in [328, p. 71] with $n = n(\xi) = ((1 + 10^{-3}\xi_3)^2 + 10^{-6}\xi_4^2)/(4\varepsilon(\xi)) + 1 \geq 1$, we find that $\|y_0(\xi)\|_{L^2(\mathcal{D})} \leq \varepsilon(\xi)$ and

$$\|y_0(\xi)\|_Y^2 \leq [\varepsilon(\xi)^2((1 + 10^{-3}\xi_3)^2 + 10^{-6}\xi_4^2) + \varepsilon(\xi)^4]^2 + \varepsilon(\xi)^2. \tag{2.7.6}$$

In light of Lemma 2.3.4, it suffices to show that each addend in (2.7.5) is uniformly integrable. Lemmas 2.3.4 and 2.9.2 ensure the uniform integrability of $(3/\kappa(\cdot))\|u\|_{H^{-1}(\mathcal{D})}$. Moreover, (2.7.6) and Lemma 2.9.2 reveal the uniform integrability of $\|y_0(\cdot)\|_Y^2$ and of $\|y_0(\cdot)\|_{L^2(\mathcal{D})}^2$. Lemma 2.3.4 further implies that the first addend, $4\|y_0(\cdot)\|_Y$, in (2.7.5) is uniformly integrable.
It must yet be shown that the third addend, $(3/\kappa(\cdot))\|y_0(\cdot)\|_Y \|y_0(\cdot)\|_{L^2(\mathcal{D})}$, in (2.7.5) is uniformly integrable. Young's inequality implies that $(3/\kappa(\cdot))^2 \|y_0(\cdot)\|_{L^2(\mathcal{D})}^2 + \|y_0(\xi)\|_Y^2$ dominates $(3/\kappa(\cdot))\|y_0(\cdot)\|_Y \|y_0(\cdot)\|_{L^2(\mathcal{D})}$. We have $(3/\kappa(\cdot))^2 \|y_0(\cdot)\|_{L^2(\mathcal{D})}^2 \leq (3/\kappa(\cdot))^2 \kappa(\cdot)^2/4 = 9/4$. Putting together the statements and using Lemma 2.3.4 once more, we conclude that $\|S(u,\cdot)\|_Y^2$ is uniformly integrable. □

If the solution of the steady Burgers' equation (2.7.2) is unique, then Lemmas 2.3.5 and 2.7.2 ensure the uniform integrability of $\widehat{J}(u,\cdot)$ for $u \in U_{\mathrm{ad}}$. In this case, Assumption 2.3.3 holds.
For $(y,u,\xi) \in Y \times U \times \mathbb{R}^p$ with $e(y,u,\xi) = 0$, the operator $e_y(y,u,\xi) \in \mathscr{L}(Y, V^* \times \mathbb{R}^2)$ is surjective [328, Thm. 3.3 (p. 81)]. We show that this operator is also injective (a fact that has also been stated in [188, p. A1866]). Combined with the bounded mapping theorem and the implicit function theorem, we conclude that the solution of the PDE (2.7.2) is locally unique.

Lemma 2.7.3. *If $(y, u, \xi) \in Y \times U \times \mathbb{R}^p$ fulfills $e(y, u, \xi) = 0$, then $e_y(y, u, \xi) \in \mathscr{L}(Y, V^* \times \mathbb{R}^2)$ is bijective, where $e : Y \times U \times \mathbb{R}^p \to V^* \times \mathbb{R}^2$ is defined in* (2.7.3).

Proof. The surjectivity of $e_{(y,u)}(y, u, \xi)$ [328, Thm. 3.3 (p. 81)] (see also [329, Prop. 3.3]) implies that of $e_y(y, u, \xi)$.
Our proof of injectivity of $e_y(y, u, \xi)$ is built on the proof of [328, Thm. 3.3 (p. 81)]. Using [328, eqns. (4.23) and (4.25)], we have, for each $\varphi \in V$ and $w \in Y$,

$$\langle \mathrm{D}_y e_1(y, u, \xi)w, \varphi\rangle_{V^*,V} = \int_{\mathcal{D}} \kappa(\xi)w'\varphi' + (yw)'\varphi \mathrm{d}x \quad \text{and} \quad \mathrm{D}_y e_2(y, u, \xi)w = (w(0), w(1)).$$

To prove the injectivity, we show that whenever $w \in Y = H^1(\mathcal{D})$ fulfills

$$\langle \mathrm{D}_y e_1(y, u, \xi)w, \varphi\rangle_{V^*,V} = 0 \quad \text{for all} \quad \varphi \in V \quad \text{and} \quad (w(0), w(1)) = 0, \tag{2.7.7}$$

it holds that $w = 0$ using the Fredholm-type alternative [196, Thm. 8.6-1]. Let $w \in Y$ satisfy (2.7.7). Then $w \in V = H^1_0(\mathcal{D})$. We define $K : V \to V^*$ and $A : V \to V^*$ by

$$\langle Kw, \varphi\rangle_{V^*,V} = \int_{\mathcal{D}} (yw)'\varphi \mathrm{d}x \quad \text{and} \quad \langle Aw, \varphi\rangle_{V^*,V} = \int_{\mathcal{D}} \kappa(\xi)w'\varphi' \mathrm{d}x \quad \text{for all} \quad \varphi \in V.$$

The operators A and K are linear and bounded, and A is boundedly invertible and K is compact [328, p. 82]. We define the compact operator $T : V \to V$ by $Tw = -A^{-1}Kw$, and $G : V \to V^*$ by $G = A + K$. The equations in (2.7.7) are equivalent to $Gw = 0$, and we have $Gw = -A(T - I)w$. In particular $(T - I)w = 0$. If $f \in V^*$ and $T^*f = f$, then $f = 0$ [328, p. 82]. Hence, the theorem on the solvability of operator equations involving compact linear operators [196, Thm. 8.5-1], implies that, for each $y \in V$, the equation $(T - I)x = y$ has a solution $x \in V$. Combining $(T - I)w = 0$ and the Fredholm-type alternative [196, Thm. 8.6-1], we find that $w = 0$. We conclude that $e_y(y, u, \xi)$ is injective. □

Discretization and Numerical Results

We transformed the steady Burgers' equation (2.7.2) to one with homogeneous boundary conditions, and discretized it using continuous piecewise linear finite elements on a uniform mesh of the domain $\mathcal{D}$ with 2000 elements as in [188, sect. 5.2.2].
We approximated the DROP (2.7.1) with the DROP (2.1.5) and used Algorithm 2 to compute a stationary point of (2.1.5). We implemented Algorithm 2 in `Python` using `UFL` [7, 5] to evaluate the derivatives of J and e, and `FEniCS` [6, 220] to compute the solutions to the PDEs (see section 2.9.3).
We chose the initial point $u_0 = 0$, and $(\tau_1, \nu_1, \eta_1) = 10^{-2}(1, 10^{-2}, 1)$ in Algorithm 2 and used the rule $(\tau_{k+1}, \nu_{k+1}, \eta_{k+1}) = 10^{-1}(\tau_k, 10^{-1}\nu_k, \eta_k)$ to update the smoothing parameters. Owing to the term $(2\nu_k)^{1/2}$ in (2.2.10), the parameter ν_k was decreased faster than τ_k and η_k. Algorithm 2 used `moola` [286] with its default settings except of using the termination tolerance 10^{-4} for each inner iteration of Algorithm 2, Wolfe line search, and L-BFGS. The TRPs (2.2.7) were solved using the Moré–Sorensen algorithm [238].
Figure 2.1 depicts the controls u^*_N and u^*_{DR} and their corresponding states for three different ambiguity sets, where u^*_N is a stationary point of the nominal control problem (2.2.2) and u^*_{DR} is the final iterate of Algorithm 2. The robust controls depicted in Figure 2.1 have a similar structure as those obtained in [191, sect. 6.2] via the minimization of the superquantile/conditional value-at-risk using the sample average approximation. For the approach in [191], each evaluation of the cost function and its gradient requires as many solutions of the Burgers' equation (2.7.2) as samples used, ranging from 19,000 to 23,000 in [191, sect. 6.2]. Our approach requires 37

Table 2.1: *Iteration history of Algorithm 2 applied to the approximated DROP of steady Burgers' equation* (2.7.2)*, with* $\Delta = \sigma_1 = 0.1$*,* $\sigma_0 = 0$*,* $\bar{\mu} = 0$*,* $\bar{\Sigma} = I$ *and* $t^k = (\tau_k, \nu_k, \eta_k)$*.*

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

Table 2.2: *Statistics (see* (2.7.8)*) for nominal control* u_N^* *and distributionally robust control* $u_{DR}^*(\Delta)$*, associated with steady Burgers' equation* (2.7.2)*, with* $\Delta = \sigma_1 = 0.1$*,* $\sigma_0 = 0$*,* $\bar{\mu} = 0$*, and* $\bar{\Sigma} = I$*.*

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

solutions of (2.7.2) and 629 solutions of linear PDEs in total for the setup displayed in Table 2.1. We note that the cost function used here and in [191, sect. 6.2] are different and, hence, corresponding optimal controls cannot be compared directly. Our point is, however, that our scheme produced meaningful controls with moderate computational costs.

We present detailed numerical results for the data $\Delta = \sigma_1 = 0.1$, $\sigma_0 = 0$, $\bar{\Sigma} = I$ and $\bar{\mu} = 0$. Table 2.1 provides an iteration history of Algorithm 2. It displays the objective function value of (2.2.1) and the U-norm of its gradient at the computed stationary point u^k of (2.2.1) for each outer iteration k of Algorithm 2. Moreover, it shows the number of inner iterations performed, a relative distance of subsequent stationary points, and the number of objective and gradient evaluations. The number of outer iterations of Algorithm 2 and the error of subsequent iterates of Algorithm 2 decrease monotonically.

For the stationary points u_N^* and u_{DR}^*, Table 2.2 displays the statistics

$$\begin{aligned} \mathrm{E}^m(u) &= \max_{1\le i\le m} \mathbb{E}_{P_i}[\widehat{J}(u,\xi)], \quad \mathrm{StD}^m(u) = \max_{1\le i\le m} \mathrm{StD}_{P_i}[\widehat{J}(u,\xi)], \\ \mathrm{ur SD}^m(u) &= \max_{1\le i\le m} \mathbb{E}_{P_i}\big[\big(\widehat{J}(u,\xi) - \mathbb{E}_{P_i}[\widehat{J}(u,\xi)]\big)_+^r\big], \quad r \in \{1,2\}, \\ Q^m_\beta(u) &= \max_{1\le i\le m} \mathrm{VaR}_{P_i,\beta}(\widehat{J}(u,\xi)), \end{aligned} \tag{2.7.8}$$

where $P_i = \mathcal{N}(\hat{\mu}_i, \hat{\sigma}_i^2 \bar{\Sigma}) \in \mathcal{P}$, $m \in \mathbb{N}$, StD_{P_i} is the standard deviation, $\mathrm{ur SD}_{P_i}$ is the upper-r-semideviation and $\mathrm{VaR}_{P_i,\beta}$, $\beta \in (0,1)$, is the value-at-risk. Here, $\hat{\mu}_i$ are uniformly and independently distributed over $\{\, \mu \in \mathbb{R}^p : \|\mu\|_2 \le \Delta \,\}$ and $\hat{\sigma}_i$ on $[\sigma_0, \sigma_1]$ for $i = 1, \ldots, m$. We chose $m = 10$ and approximated the quantities in (2.7.8) with 1000 independent samples. The statistics reported in Table 2.2 verify empirically that the distributionally robust control is more robust than the nominal control. We obtained similar results as in Table 2.2 for different choices of the parameters Δ, σ_0 and σ_1. The numerical results indicate that the objective function F defined in (2.1.6) may be nonsmooth at $u_{\mathrm{DR}}^*(\Delta)$ for several different values of Δ.

2.7.2 DRO of Unsteady Burgers' Equation

We consider

$$\min_{u\in U_{\mathrm{ad}}} \sup_{P\in\mathcal{P}} \big\{\, \mathbb{E}_P[(1/2)\|S(u,\xi) - y_d\|_H^2] + (\alpha/2)\|u_1\|^2_{L^2(\mathcal{I})} + (\alpha/2)\|u_2\|^2_{L^2(\mathcal{I})} \big\}, \tag{2.7.9}$$

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

[4]Here $W(\mathcal{I}; L^2(\mathcal{D}), H^1(\mathcal{D})) = \{\, v \in L^2(\mathcal{I}, H^1(\mathcal{D})) : v_t \in L^2(\mathcal{I}, H^1(\mathcal{D})^*) \,\}$ is equipped with the norm $\|\cdot\|_Y = (\|\cdot\|^2_{L^2(\mathcal{I};H^1(\mathcal{D}))} + \|\cdot_t\|^2_{L^2(\mathcal{I};H^1(\mathcal{D})^*)})^{1/2}$ [151, pp. 39–40].

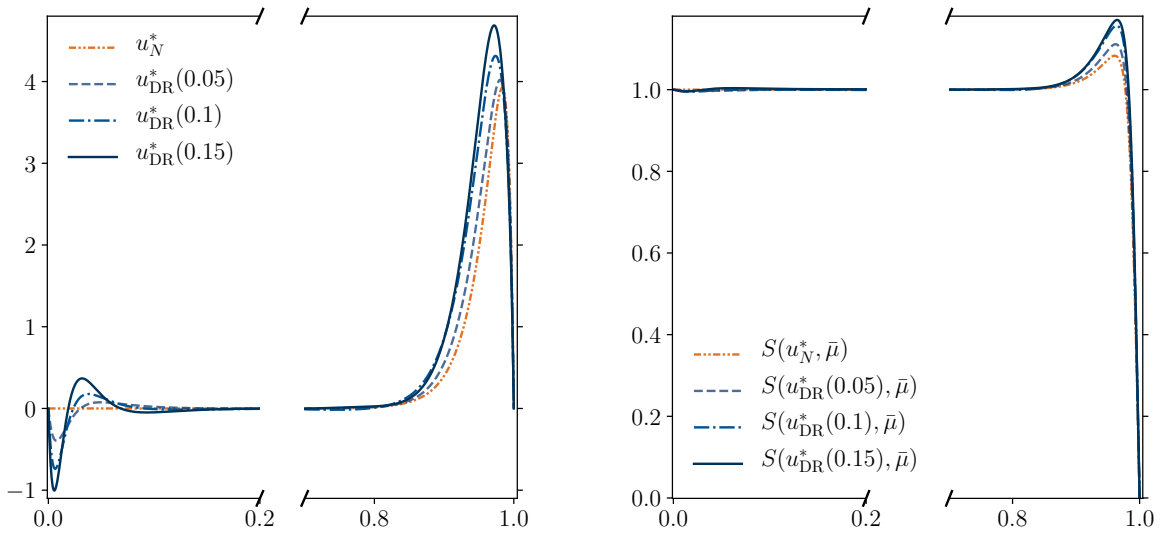


FIGURE 2.1: *(Left) Stationary control* u_N^* *of* (2.2.2) *and* $u_{DR}^*(\Delta)$ *of* (2.1.5) *for* $\Delta = \sigma_1 = 0.05$, $\Delta = \sigma_1 = 0.1$ *and* $\Delta = \sigma_1 = 0.15$, *associated with the approximated DROP of the steady Burgers' equation* (2.7.2). *The remaining data that defines the ambiguity set* $\mathcal{P}$ *(see* (2.1.2)*) is* $\sigma_0 = 0$, $\bar{\mu} = 0$, $\bar{\Sigma} = I$. *(Right) Corresponding states evaluated at* $(u_N^*, \bar{\mu})$ *and* $(u_{DR}^*(\Delta), \bar{\mu})$. *Each graph is approximately constant between the breaks.*

Burgers' equation

$$\begin{aligned} &y_t(x,t) = \kappa(\xi) y_{xx}(x,t) - y(x,t) y_x(x,t) + (\xi_4/100)t, \quad x \in \mathcal{D}, \\ &y(x,0) = \phi(x,\xi), \quad x \in \mathcal{D}, \qquad y_x(0,t) = u_1(t), \quad y_x(1,t) = u_2(t), \quad t \in \mathcal{I}, \end{aligned} \tag{2.7.10}$$

where $(u_1, u_2) \in U$, $p = 4$, $\kappa : \mathbb{R}^p \to \mathbb{R}_{++}$ is given by $\kappa(\xi) = 10^{\xi_1 - 1}$, and $\phi : \mathcal{D} \times \mathbb{R}^p \to \mathbb{R}$ is defined by $\phi(x,\xi) = (1 - 10\xi_2)x^2(1 + 10\xi_3 - x)(1 - x)$. The parameterized model (2.7.10) is based on that used in [62, sect. 7]. We equip $U = L^2(\mathcal{I}) \times L^2(\mathcal{I})$ with the norm $\|\cdot\|_U = (\|\cdot\|_{L^2(\mathcal{I})}^2 + \|\cdot\|_{L^2(\mathcal{I})}^2)^{1/2}$. Deterministic optimal control problems subject to (2.7.10) are considered in, for example, [328, 332, 330]. The sensitivity analysis performed in [62, sect. 7] of the optimal state and controls of a deterministic optimal control problem subject to (2.7.10) revealed the sensitivity of the computed control w.r.t. infinitesimal perturbations of ξ about its nominal value $\bar{\mu} = 0$.
We show that Assumptions 2.3.1, 2.3.2 and 2.3.12 hold. We define $V = L^2(\mathcal{I}; H^1(\mathcal{D}))$, and identify with $L^2(\mathcal{I}; H^1(\mathcal{D})^*)$ the dual of V as in [328, p. 145]. Moreover, we define $e = (e_1, e_2) : Y \times U \times \mathbb{R}^p \to V^* \times L^2(\mathcal{D})$ by $e_2(y,u,\xi) = y(\cdot,0) - \phi(\cdot,\xi)$ and by

$$\begin{aligned} \langle e_1(y,u,\xi), v\rangle_{V^*,V} = \langle y_t, v\rangle_{V^*,V} &+ \int_{\mathcal{I}} [u_1(t)v(0,t) - u_2(t)v(1,t)]\mathrm{d}t \\ &+ \int_{\mathcal{I}}\int_{\mathcal{D}} [\kappa(\xi) y_x(x,t) v_x(x,t) + (y(x,t) y_x(x,t) - 10^{-2}\xi_4 t) v(x,t)]\mathrm{d}x\mathrm{d}t, \end{aligned} \tag{2.7.11}$$

for all $v \in V$; cf. [328, p. 145].

Lemma 2.7.4. *The operator* e *defined in* (2.7.11) *is twice continuously differentiable.*

Proof. For fixed $\xi \in \mathbb{R}^p$, the derivations in [328, p. 146] imply that $e(\cdot,\cdot,\xi)$ is twice continuously differentiable. We fix $(y,u,\xi) \in Y \times U \times \mathbb{R}^p$. Since ϕ is a polynomial and $\mathcal{D}$ is bounded, the operator $e_2(y,u,\cdot) : \mathbb{R}^p \to L^2(\mathcal{D})$ is twice continuously differentiable. For each $v \in V$ and s,

$h \in \mathbb{R}^p$, we have

$$\begin{aligned}\langle \mathrm{D}_\xi e_1(y,u,\xi)h, v\rangle_{V^*,V} &= \int_{\mathcal{I}}\int_{\mathcal{D}} [\ln(10)10^{\xi_1-1}h_1 y_x v_x - 10^{-2}h_4 tv]\mathrm{d}x\mathrm{d}t,\\ \langle \mathrm{D}_{\xi\xi} e_1(y,u,\xi)[h,s], v\rangle_{V^*,V} &= \int_{\mathcal{I}}\int_{\mathcal{D}} [\ln(10)^2 10^{\xi_1-1}h_1 s_1 y_x v_x]\mathrm{d}x\mathrm{d}t.\end{aligned} \tag{2.7.12}$$

We only verify the first formula. The latter identity reveals the continuity of the second derivative of $\mathrm{D}_{\xi\xi}e_1$. Using the Cauchy–Schwarz inequality and the continuous embedding $V \hookrightarrow H$, we find that, for all $v \in V$ with $\|v\|_V \le 1$,

$$|\langle e_1(y,u,\xi+h) - e_1(y,u,\xi) - \mathrm{D}_\xi e_1(y,u,\xi)h, v\rangle_{V^*,V}| \le |\kappa(\xi+h) - \kappa(\xi) - \mathrm{D}\kappa(\xi)h|\|y\|_Y.$$

Combining the pieces with the calculus rules [97, Thms. 8.9.1 and 8.12.6] yields the claims. □

The tracking-type cost function $J : Y \times U \times \mathbb{R}^p \to \mathbb{R}_+$ defined by $J(y,u,\xi) = (1/2)\|y - y_d\|_H^2 + (\alpha/2)\|u_1\|_{L^2(\mathcal{I})}^2 + (\alpha/2)\|u_2\|_{L^2(\mathcal{I})}^2$ is convex and infinitely many times continuously differentiable, and e is twice continuously differentiable (see Lemma 2.7.4). Hence Assumption 2.3.2 and Assumption 2.3.12 (a) hold.
For each fixed $(u,\xi) \in U \times \mathbb{R}^p$, the state equation $e(y,u,\xi) = 0$ has a unique solution $S(u,\xi)$ [331, Thm. 2.3], and $e_y(S(u,\xi),u,\xi) \in \mathscr{L}(Y,Z)$ is bijective [331, Prop. 2.5]. The bounded mapping theorem implies that $e_y(S(u,\xi),u,\xi)$ is boundedly invertible. Hence Assumption 2.3.12 (b) holds. The continuity of $S(u,\cdot)$ follows from the implicit function theorem. Combined with Lemma 2.7.5, we conclude that the conditions of Assumption 2.3.1 are fulfilled.

Lemma 2.7.5. *For all $\xi \in \mathbb{R}^p$, the parameterized solution operator $S(\cdot,\xi) : U \to Y$ for the unsteady Burgers' equation* (2.7.10) *is weakly-weakly continuous.*

Proof. We verify the hypotheses of Proposition 2.9.3. Fix $\xi \in \mathbb{R}^p$. The proof of [331, Thm. 2.4] implies that $\{\,(y,u) \in Y \times U : e(y,u,\xi) = 0\,\}$ is weakly sequentially closed. From [331, Thm. 2.3] and its proof, we deduce the existence of a constant $C(\xi) \in \mathbb{R}_{++}$ such that

$$\|S(u,\xi)\|_Y \le C(\xi)(1 + \|u_1\|_{L^2(\mathcal{I})} + \|u_2\|_{L^2(\mathcal{I})}) \quad \text{for all} \quad u = (u_1,u_2) \in U \tag{2.7.13}$$

(see also [328, pp. 141–142]). For each $(u,\xi) \in U \times \mathbb{R}^p$, the unsteady Burgers' equation (2.7.11) has a unique solution. The norms $\|\cdot\|_U = (\|\cdot\|_{L^2(\mathcal{I})}^2 + \|\cdot\|_{L^2(\mathcal{I})}^2)^{1/2}$ and $\|\cdot\|_{L^2(\mathcal{I})} + \|\cdot\|_{L^2(\mathcal{I})}$ are equivalent. Moreover, $Y = W(\mathcal{I}; L^2(\mathcal{D}), H^1(\mathcal{D}))$ is a Hilbert space [151, Thm. 1.32] and, hence, it is reflexive. Thus Proposition 2.9.3 implies that $S(\cdot,\xi)$ is weakly-weakly continuous. □

We show that Assumption 2.3.3 is fulfilled by verifying the hypotheses of Lemma 2.3.5.

Lemma 2.7.6. *For $u \in U$, the function $\|S(u,\cdot)\|_H^2$ is uniformly integrable.*

Proof. Since $V = L^2(\mathcal{I}; H^1(\mathcal{D}))$ is separable [159, Prop. 1.2.29], $S(u,\cdot) : \mathbb{R}^p \to Y$ is continuous, and $Y \hookrightarrow V \hookrightarrow H$ are continuous, the mappings $\|S(u,\cdot)\|_H$ and $\|S(u,\cdot)\|_V$ are measurable.
Owing to Lemma 2.3.4 and $\|S(u,\cdot)\|_H^2 \le \|S(u,\cdot)\|_V^2$, it suffices to prove that $\|S(u,\cdot)\|_V^2$ is uniformly integrable. The proof of [328, Thm. 4.2 (p. 141)] and of [331, Prop. A.6] imply the stability estimate

$$\|S(u,\xi)\|_V^2 \le \frac{3|\xi_4|}{\kappa(\xi)^2} + \frac{\|\phi(\cdot,\xi)\|_{L^2(\mathcal{D})}^2}{\kappa(\xi)} + 6\|u_1\|_{L^2(\mathcal{I})}^2 + 6\|u_2\|_{L^2(\mathcal{I})}^2 + \frac{c_1 c_2(\xi)^2}{\kappa(\xi)^4} + 2c_2(\xi), \tag{2.7.14}$$

where $c_1 > 0$ and $c_2 : \mathbb{R}^p \to \mathbb{R}_+$ is defined by $c_2(\xi) = 4(1 + \|\phi(\cdot,\xi)\|_{L^2(\mathcal{D})}^2)^2$. We have $\|\phi(\cdot,\xi)\|_{L^2(\mathcal{D})} = 630^{-1}(1 - 10\xi_2)^2(600\xi_3^2 + 45\xi_3 + 1)$. Lemmas 2.3.4, 2.9.1 and 2.9.2, imply that each addend in (2.7.14) is uniformly integrable. Hence $\|S(u,\cdot)\|_V^2$ is uniformly integrable. □

Lemmas 2.3.5 and 2.7.6 imply that Assumption 2.3.3 holds. To establish Assumption 2.3.12 (c), we apply Lemma 2.7.7. The embedding $Y \hookrightarrow H = L^2(\mathcal{I}; L^2(\mathcal{D}))$ is compact [312, Thm. 2.1 (p. 271)], and Y is a Hilbert space [151, Thm. 1.32] and, hence, it is reflexive. Combined with the assertions of Lemmas 2.3.16 and 2.7.7, we conclude that Assumption 2.3.12 (c) holds.

Lemma 2.7.7. *For all $s_\xi \in \mathbb{R}^p$, $S_\xi(\cdot, \bar{\mu})s_\xi : U \to Y$ and $S_{\xi\xi}(\cdot, \bar{\mu})[s_\xi, s_\xi] : U \to Y$ are weakly-weakly continuous. Here, $S : U \times \mathbb{R}^p \to Y$ is the solution operator for* (2.7.10).

Lemma 2.7.7 is established using Lemmas 2.7.8 and 2.7.9.

Lemma 2.7.8. *There exists a constant $c > 0$ such that, for all v, $w \in Y$ and $\varphi \in V$,*

$$\left| \int_{\mathcal{I}} \int_{\mathcal{D}} v_x w \varphi \mathrm{d}x\mathrm{d}t \right| \leq c \|v\|_{L^2(\mathcal{I}; H^1(\mathcal{D}))} \|w\|_{C(\bar{\mathcal{I}}; L^2(\mathcal{D}))} \|\varphi\|_V. \tag{2.7.15}$$

Moreover, if (v_k), $(w_k) \subset Y$ satisfy $v_k \rightharpoonup v \in Y$ and $w_k \rightharpoonup w \in Y$, then $\int_{\mathcal{I}} \int_{\mathcal{D}} (v_k w_k)_x \varphi \mathrm{d}x\mathrm{d}t \to \int_{\mathcal{I}} \int_{\mathcal{D}} (vw)_x \varphi \mathrm{d}x\mathrm{d}t$ as $k \to \infty$ for all $\varphi \in V$.

Proof. The embeddings $H^1(\mathcal{D}) \hookrightarrow C(\bar{\mathcal{D}})$ and $Y \hookrightarrow C(\bar{\mathcal{I}}; L^2(\mathcal{D}))$ are continuous [151, Thms. 1.14 and 1.32]. Combined with Hölder's inequality, we find that, for some $c > 0$ and all v, $w \in Y$ and $\varphi \in V$,

$$\begin{aligned} \left| \int_{\mathcal{I}} \int_{\mathcal{D}} v_x w \varphi \mathrm{d}x\mathrm{d}t \right| &\leq \int_{\mathcal{I}} \|\varphi(t)\|_{C(\bar{\mathcal{D}})} \|v(t)\|_{H^1(\mathcal{D})} \|w(t)\|_{L^2(\mathcal{D})} \mathrm{d}t & (2.7.16) \\ &\leq c \|v\|_{L^2(\mathcal{I}; H^1(\mathcal{D}))} \|w\|_{C(\bar{\mathcal{I}}; L^2(\mathcal{D}))} \|\varphi\|_V. & (2.7.17) \end{aligned}$$

The compact embedding $Y \hookrightarrow C(\bar{\mathcal{I}}; L^2(\mathcal{D}))$ [312, Thm. 2.1 (p. 271)], the stability estimate (2.7.15), and the identity $\int_{\mathcal{I}} \int_{\mathcal{D}} (vw)_x \varphi \mathrm{d}x\mathrm{d}t = \int_{\mathcal{I}} \int_{\mathcal{D}} [v_x w \varphi + v w_x \varphi] \mathrm{d}x\mathrm{d}t$, valid for all v, $w \in Y$ and $\varphi \in V$, imply the second assertion. □

Lemma 2.7.9. *For e defined in* (2.7.11)*, there exists a function $\rho : \mathbb{R}^{3+p} \to \mathbb{R}$ such that*

$$\|e_y(S(u, \xi), u, \xi)^{-1}(g, h)\|_Y \leq \rho(\|S(u, \xi)\|_Y, \|g\|_{V^*}, \|h\|_{L^2(\mathcal{D})}, \xi)$$

for all $(u, g, h, \xi) \in U \times V^* \times L^2(\mathcal{D}) \times \mathbb{R}^p$ and, for each $\xi \in \mathbb{R}^p$, $\rho(\cdot, \xi)$ is monotonically increasing.

Proof. The proof is inspired by that of [331, Prop. 2.5] (see also [321, Prop. 10.4]). Fix $(u, g, h, \xi) \in U \times V^* \times L^2(\mathcal{D}) \times \mathbb{R}^p$. Since $e_y(S(u, \xi), u, \xi)$ is boundedly invertible [331, Prop. 5.2], there exists a unique $w \in Y$ with $e_y(S(u, \xi), u, \xi)w = (g, h)$, which is equivalent to $w(0) = h$ and

$$\langle w_t(t), \varphi \rangle_{H^1(\mathcal{D})^*, H^1(\mathcal{D})} + \int_{\mathcal{D}} [\kappa(\xi) w_x(t) \varphi' + (S(u, \xi)w)_x(t) \varphi] \mathrm{d}x = \langle g(t), \varphi \rangle_{H^1(\mathcal{D})^*, H^1(\mathcal{D})} \tag{2.7.18}$$

for all $\varphi \in H^1(\mathcal{D})$ and for almost every $t \in \mathcal{I}$; see [331, p. 254]. We fix $t \in \mathcal{I}$ and choose $\varphi = w(t)$ in (2.7.18). Using $\langle w_t(t), w(t) \rangle_{H^1(\mathcal{D})^*, H^1(\mathcal{D})} = (1/2)\mathrm{d}/\mathrm{d}t \|w(t)\|^2_{L^2(\mathcal{D})}$ [297, Prop. 1.2 (p. 106)] (see also [312, Lem. 1.2 (pp. 260–161)]) and Hölder's inequality, we find that

$$\begin{aligned} \frac{1}{2} \frac{\mathrm{d}}{\mathrm{d}t} \|w(t)\|^2_{L^2(\mathcal{D})} + \kappa(\xi) \|w(t)\|^2_{H^1(\mathcal{D})} &\leq \|g(t)\|_{H^1(\mathcal{D})^*} \|w(t)\|_{H^1(\mathcal{D})} + \kappa(\xi) \|w(t)\|_{L^2(\mathcal{D})} \\ &\quad + \|S(u, \xi)(t)\|_{H^1(\mathcal{D})} \|w(t)\|_{L^2(\mathcal{D})} \|w(t)\|_{C(\bar{\mathcal{D}})} \\ &\quad + \|S(u, \xi)(t)\|_{L^2(\mathcal{D})} \|w(t)\|_{H^1(\mathcal{D})} \|w(t)\|_{L^\infty(\mathcal{D})}. \end{aligned}$$

Agmond's inequality ensures $\|w(t)\|_{L^\infty(\mathcal{D})} \leq C\|w(t)\|^{1/2}_{L^2(\mathcal{D})}\|w(t)\|^{1/2}_{H^1(\mathcal{D})}$ for some absolute constant $C \in \mathbb{R}_{++}$ [313, eq. (2.21)]. Combined with Young's inequality and the continuous embedding $H^1(\mathcal{D}) \hookrightarrow C(\bar{\mathcal{D}})$ [151, Thm. 1.14], we deduce the existence of c_1, $c_2 \in \mathbb{R}_{++}$ such that

$$\begin{aligned}\tfrac{1}{2}\tfrac{\mathrm{d}}{\mathrm{d}t}\|w(t)\|^2_{L^2(\mathcal{D})} + \tfrac{\kappa(\xi)}{2}\|w(t)\|^2_{H^1(\mathcal{D})} &\leq \tfrac{c_1}{\kappa(\xi)}\|g(t)\|^2_{H^1(\mathcal{D})^*} \\ &\quad + c_2\big(\kappa(\xi) + \tfrac{1}{\kappa(\xi)}\|S(u,\xi)(t)\|^2_{H^1(\mathcal{D})} + \tfrac{1}{\kappa(\xi)}\|S(u,\xi)(t)\|^4_{L^2(\mathcal{D})}\big)\|w(t)\|^2_{L^2(\mathcal{D})}.\end{aligned}$$

Integrating over $(0,T)$ for fixed $T \in (0,1)$, and using $w(0) = h$ and the fact that the dual of $V = L^2(\mathcal{I}; H^1(\mathcal{D}))$ is identified with $L^2(\mathcal{I}; H^1(\mathcal{D})^*)$, we find that

$$\begin{aligned}&\|w(T)\|^2_{L^2(\mathcal{D})} + \kappa(\xi)\int_0^T \|w(t)\|^2_{H^1(\mathcal{D})}\mathrm{d}t \leq \tfrac{2c_1}{\kappa(\xi)}\|g\|^2_{V^*} + \|h\|^2_{L^2(\mathcal{D})} \\ &+ 2c_2\int_0^T \big(\kappa(\xi) + \tfrac{1}{\kappa(\xi)}\|S(u,\xi)(t)\|^2_{H^1(\mathcal{D})} + \tfrac{1}{\kappa(\xi)}\|S(u,\xi)(t)\|^4_{L^2(\mathcal{D})}\big)\|w(t)\|^2_{L^2(\mathcal{D})}\mathrm{d}t.\end{aligned} \tag{2.7.19}$$

Combined with the continuous embedding $Y \hookrightarrow C(\bar{\mathcal{I}}; L^2(\mathcal{D}))$ [297, Prop. 1.2 (p. 106)], Gronwall's inequality (see, e.g., [338, p. 317]) ensures, for all $t \in \bar{\mathcal{I}}$,

$$\|w(t)\|^2_{L^2(\mathcal{D})} \leq \big(\tfrac{2c_1}{\kappa(\xi)}\|g\|^2_{V^*} + \|h\|^2_{L^2(\mathcal{D})}\big)\exp\big(2c_2\kappa(\xi) + \tfrac{2c_2}{\kappa(\xi)}\|S(u,\xi)\|^2_V + \tfrac{2c_2}{\kappa(\xi)}\|S(u,\xi)\|^4_{L^4(\mathcal{I};L^2(\mathcal{D}))}\big)$$

(see also [331, eqns. (2.7) and (2.8)]). Using (2.7.19), and the continuity of $Y \hookrightarrow L^4(\mathcal{I}; L^2(\mathcal{D}))$ and of $Y \hookrightarrow V$, we deduce the existence of a function $\rho_1 : \mathbb{R}^3 \times \mathbb{R}^p \to \mathbb{R}_{++}$ such that $\|w\|_V \leq \rho_1(\|S(u,\xi)\|_Y, \|g\|_{V^*}, \|h\|_{L^2(\mathcal{D})}, \xi)$, and $\rho_1(\cdot,\xi)$ is increasing for all $\xi \in \mathbb{R}^p$.[5]
Next, we derive an upper bound on $\|w_t\|_{L^2(\mathcal{D};H^1(\mathcal{D})^*)}$. From (2.7.18) and Hölder's inequality, we obtain, for all $\varphi \in H^1(\mathcal{D})$,

$$\begin{aligned}|\langle w_t(t), \varphi\rangle_{H^1(\mathcal{D})^*,H^1(\mathcal{D})}| &\leq \|g(t)\|_{H^1(\mathcal{D})^*}\|\varphi\|_{H^1(\mathcal{D})} + \kappa(\xi)\|w(t)\|_{H^1(\mathcal{D})}\|\varphi\|_{H^1(\mathcal{D})} \\ &\quad + \|S(u,\xi)(t)\|_{H^1(\mathcal{D})}\|w(t)\|_{L^2(\mathcal{D})}\|\varphi\|_{C(\bar{\mathcal{D}})} \\ &\quad + \|S(u,\xi)(t)\|_{L^2(\mathcal{D})}\|w(t)\|_{H^1(\mathcal{D})}\|\varphi\|_{C(\bar{\mathcal{D}})}.\end{aligned}$$

Combined with Jensen's inequality and the continuous embedding $H^1(\mathcal{D}) \hookrightarrow C(\bar{\mathcal{D}})$, we deduce the existence of a constant $c_3 \in \mathbb{R}_{++}$ such that

$$\begin{aligned}\|w_t\|^2_{L^2(\mathcal{D};H^1(\mathcal{D})^*)} &\leq c_3\|g\|^2_{V^*} + c_3\kappa(\xi)\|w\|^2_{L^2(\mathcal{I};L^2(\mathcal{D}))} \\ &\quad + c_3\|w\|^2_{C(\bar{\mathcal{I}};L^2(\mathcal{D}))}\|S(u,\xi)\|^2_V + c_3\|S(u,\xi)\|^2_{C(\bar{\mathcal{I}};L^2(\mathcal{D}))}\|w\|^2_V.\end{aligned}$$

Putting together the pieces, we deduce the existence of $\rho : \mathbb{R}^3 \times \mathbb{R}^p \to \mathbb{R}_{++}$ such that $\|w\|_Y \leq \rho(\|S(u,\xi)\|_Y, \|g\|_{V^*}, \|h\|_{L^2(\mathcal{D})}, \xi)$ and, for each $\xi \in \mathbb{R}^p$, $\rho(\cdot,\xi)$ is monotonically increasing. □

Proof of Lemma 2.7.7. Throughout the proof, we use the fact that the notions of weak and weak-star convergence in the dual space of a reflexive Banach space coincide.
We fix $(\xi, s_\xi) \in \mathbb{R}^p \times \mathbb{R}^p$, and show that $S_\xi(\cdot;\xi)s_\xi$ is weakly-weakly continuous using Proposition 2.9.3. We define $h : U \to L^2(\mathcal{D})$ and $g : U \to V^*$ by $h = \phi_\xi(\cdot,\xi)s_\xi \in L^2(\mathcal{D})$ and $g = -e_\xi(S(u,\xi),u,\xi)s_\xi \in V^*$, respectively. We fix $u \in U$. The operator $w = S_\xi(u,\xi)s_\xi$ is the unique solution to $e_y(S(u,\xi),u,\xi)w = (g(u),h(u))$. Using (2.7.12), we find that $\|g\|_{V^*} \leq \ln(10)^{\xi_1-1}|(s_\xi)_1|\|S(u,\xi)\|_Y + 10^{-2}|(s_\xi)_4|$. Combined with the stability estimate (2.7.13) and Lemma 2.7.9, we conclude that there exists a monotonically increasing function $\rho(\cdot;\xi) : \mathbb{R} \to \mathbb{R}$

[5] Since $Y \hookrightarrow C(\bar{\mathcal{I}}; L^2(\mathcal{D}))$ is continuous [297, Prop. 1.2 (p. 106)], we have, for some $C \in \mathbb{R}_{++}$ and all $v \in Y$, $(\int_{\mathcal{I}}\|v(t)\|^4_{L^2(\mathcal{D})}\mathrm{d}t)^{1/4} \leq |\mathcal{I}|\|v\|_{C(\bar{\mathcal{I}};L^2(\mathcal{D}))} = \|v\|_{C(\bar{\mathcal{I}};L^2(\mathcal{D}))} \leq C\|v\|_Y$.

such that $\|S_\xi(u,\xi)s_\xi\|_Y \leq \rho(\|u\|_U;\xi)$ for all $u \in U$. Next, we prove that $A = \{\,(w,u) \in Y \times U : e_y(S(u,\xi),u,\xi)w = (g(u),h(u))\,\}$ is weakly sequentially closed. Let $(w_k,u_k)_{\mathbb{N}_0} \subset A$ fulfill $(w_k,u_k) \rightharpoonup (w,u) \in Y \times U$ as $k \to \infty$. According to Lemma 2.7.5, $S(\cdot,\xi)$ is weakly-weakly continuous. Hence $S(u_k,\xi) \rightharpoonup S(u,\xi) \in Y$ as $k \to \infty$. Using (2.7.12) and $(y_x)_k \rightharpoonup y_x \in L^2(\mathcal{I};L^2(\mathcal{D}))$ as $k \to \infty$ [312, p. 272], we obtain that $\mathrm{D}_\xi e_1(S(u_k,\xi),u_k,\xi)s_\xi \rightharpoonup \mathrm{D}_\xi e_1(S(u,\xi),u,\xi)s_\xi$ as $k \to \infty$. Since $\mathrm{D}_\xi e_2(S(u,\xi),u,\xi) = -\mathrm{D}_\xi\phi(\cdot,\xi)$, we get $\mathrm{D}_\xi e_2(S(u_k,\xi),u_k,\xi)s_\xi \rightharpoonup \mathrm{D}_\xi e_2(S(u,\xi),u,\xi)s_\xi$ as $k \to \infty$. It must yet be shown that $\mathrm{D}_y e(S(u_k,\xi),u_k,\xi)w_k \rightharpoonup \mathrm{D}_y e(S(u,\xi),u,\xi)w$ as $k \to \infty$. We have $\mathrm{D}_y e_2(y,u,\xi)w = w(0)$ and

$$\langle \mathrm{D}_y e_1(y,u,\xi)w,\varphi\rangle_{V^*,V} = \langle w_t,\varphi\rangle_{V^*,V} + \int_{\mathcal{I}}\int_{\mathcal{D}} [\kappa(\xi)w_x\varphi_x + (S(u,\xi)w)_x\varphi]\mathrm{d}x\mathrm{d}t$$

for all $(y,u,\xi,w,\varphi) \in Y \times U \times \mathbb{R}^p \times Y \times V$ [328, p. 146]. Since $Y \hookrightarrow C(\bar{\mathcal{I}};L^2(\mathcal{D}))$ is compact, we have $w_k(0) \to w(0) \in L^2(\mathcal{D})$ as $k \to \infty$. Moreover, $w_k \rightharpoonup w \in Y$ as $k \to \infty$ implies $w_k \rightharpoonup w \in V$ and $w_k' \rightharpoonup w' \in V^*$ as $k \to \infty$ [312, p. 272]. Combined with Lemma 2.7.8, we obtain $\mathrm{D}_y e(S(u_k,\xi),u_k,\xi)w_k \rightharpoonup \mathrm{D}_y e(S(u,\xi),u,\xi)w$ as $k \to \infty$. Putting together the pieces, we find that the set A is weakly sequentially closed. Proposition 2.9.3 implies the weak-weak continuity of $S_\xi(\cdot,\xi)s_\xi$.
We show that $S_{\xi\xi}(\cdot,\xi)[s_\xi,s_\xi]$ is weakly-weakly continuous using Proposition 2.9.3. Fix $u \in U$. We define $h : U \to \mathbb{R}$ by $h(u) = \phi_{\xi\xi}[s_\xi,s_\xi]$ and $g : U \to V^*$ by

$$\begin{aligned} g(u) = &-\mathrm{D}_{\xi\xi}e_1(S(u,\xi),u,\xi)[s_\xi,s_\xi] - \mathrm{D}_{yy}e_1(S(u,\xi),u,\xi)[S_\xi(u,\xi)s_\xi,S_\xi(u,\xi)s_\xi] \\ &- 2\mathrm{D}_{y\xi}e_1(S(u,\xi),u,\xi)[S_\xi(u,\xi)s_\xi,s_\xi]. \end{aligned}$$

The operator $s = S_{\xi\xi}(u,\xi)[s_\xi,s_\xi]$ is the unique solution to $e_y(S(u,\xi),u,\xi)s = (g(u),h(u))$. Using (2.7.12) and the derivations in [328, p. 146], we find that, for all $(y,v,w,d,\varphi) \in Y^3 \times \mathbb{R}^p \times V$,

$$\begin{aligned} \langle \mathrm{D}_{yy}e_1(y,u,\xi)[v,w],\varphi\rangle_{V^*,V} &= \int_{\mathcal{I}}\int_{\mathcal{D}} (vw)_x\varphi\mathrm{d}x\mathrm{d}t, \\ \langle \mathrm{D}_{y\xi}e_1(y,u,\xi)[v,d],\varphi\rangle_{V^*,V} &= \int_{\mathcal{I}}\int_{\mathcal{D}} \ln(10)10^{\xi_1-1}d_1v_x\varphi_x\mathrm{d}x\mathrm{d}t. \end{aligned}$$

Combined with (2.7.15), Lemma 2.7.9 and the above estimates, we deduce the existence of an increasing function $\zeta(\cdot;\xi) : \mathbb{R} \to \mathbb{R}$ such that $\|S_{\xi\xi}(u,\xi)[s_\xi,s_\xi]\|_Y \leq \zeta(\|u\|_U;\xi)$ for all $u \in U$.
Next, we prove that $B = \{\,(s,u) \in Y \times U : e_y(S(u,\xi),u,\xi)s = (g(u),h(u))\,\}$ is weakly sequentially closed. Let $(s_k,u_k)_{\mathbb{N}_0} \subset B$ such that $(s_k,u_k) \rightharpoonup (s,u) \in Y \times U$ as $k \to \infty$. Lemma 2.7.8, $(s_k,u_k) \rightharpoonup (s,u) \in Y \times U$ as $k \to \infty$, and the weak-weak continuity of $S(\cdot,\xi)$ and $S_\xi(\cdot,\xi)s_\xi$ imply that $g(u_k) \rightharpoonup g(u)$ as $k \to \infty$. Lemma 2.7.8 further yields $e_y(S(u_k,\xi),u_k,\xi)s_k \rightharpoonup e_y(S(u,\xi),u,\xi)s$ as $k \to \infty$. Hence B is weakly sequentially closed. Proposition 2.9.3 implies the weak-weak continuity of $S_{\xi\xi}(\cdot,\xi)[s_\xi,s_\xi]$. □

Discretization and Numerical Results

We discretized the unsteady Burgers' equation (2.7.10) in time, using the implicit Euler scheme on a uniform mesh of the time interval $\mathcal{I} = (0,1)$ with 100 time steps as in [328, p. 155]. For the spatial discretization, we used piecewise linear finite elements on a uniform mesh of the computational domain $\mathcal{D} = (0,1)$ with 100 elements. We used $U_{\mathrm{ad}} = U$, approximated the DROP (2.7.9) by the DROP (2.1.5) and applied Algorithm 2 to (2.1.5). We chose $u^0 = u_N^*$, the stationary control of (2.2.2), and the same initial smoothing parameters and update rule as in section 2.7.1. In Algorithm 2, we used `SciPy` [327] with L-BFGS with termination tolerance $< 10^{-2}$ for each inner iteration, and terminated Algorithm 2 when $\eta_k < 10^{-4}$.

Table 2.3: *Iteration history of Algorithm 2 applied to the approximated DROP of the unsteady Burgers' equation* (2.7.10)*, with for* $\Delta = 0.1$, $\sigma_0 = 0$, $\sigma_1 = 0.01$, $\bar{\Sigma} = I$ *and* $t^k = (\tau_k, \nu_k, \eta_k)$.

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

Table 2.4: *Statistics (see* (2.7.8)*) for nominal control* u_N^* *and distributionally robust control* $u_{DR}^*(\Delta)$*, associated with the unsteady Burgers' equation* (2.7.10)*, with* $\Delta = 0.1$, $\sigma_0 = 0$, $\sigma_1 = 0.01$, $\bar{\mu} = 0$, *and* $\bar{\Sigma} = I$.

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

Figure 2.2 depicts the stationary controls u_N^* of the nominal problem (2.2.2), and the distributionally robust controls u_{DR}^* of the approximated DROP of the unsteady Burgers' equation (2.7.10) for three ambiguity sets. Whereas the nominal control has a symmetric pattern, the robust controls are asymmetric, a result of the non-symmetry of the parameterized initial condition ϕ (see (2.7.10)). The robust controls differ significantly from the nominal one. The statistics (see (2.7.8)) reported in Table 2.4 verify empirically that the distributionally robust control is more robust than the nominal one. We obtained similar numerical values for different choices of the data defining the ambiguity set. Table 2.3 provides an iteration history of Algorithm 2 applied to the approximated DROP of the unsteady Burgers' equation (2.7.10). The difference of successive iterates of the homotopy method converge to zero. The number of objective and gradient evaluations is quite low, and Algorithm 2 required only 53 solutions of the unsteady Burgers' equation (2.7.10) in total. Our numerical results indicate that the objective function F (see (2.1.6)) may be nonsmooth at $u_{\mathrm{DR}}^*(\Delta)$ for multiple values of Δ.

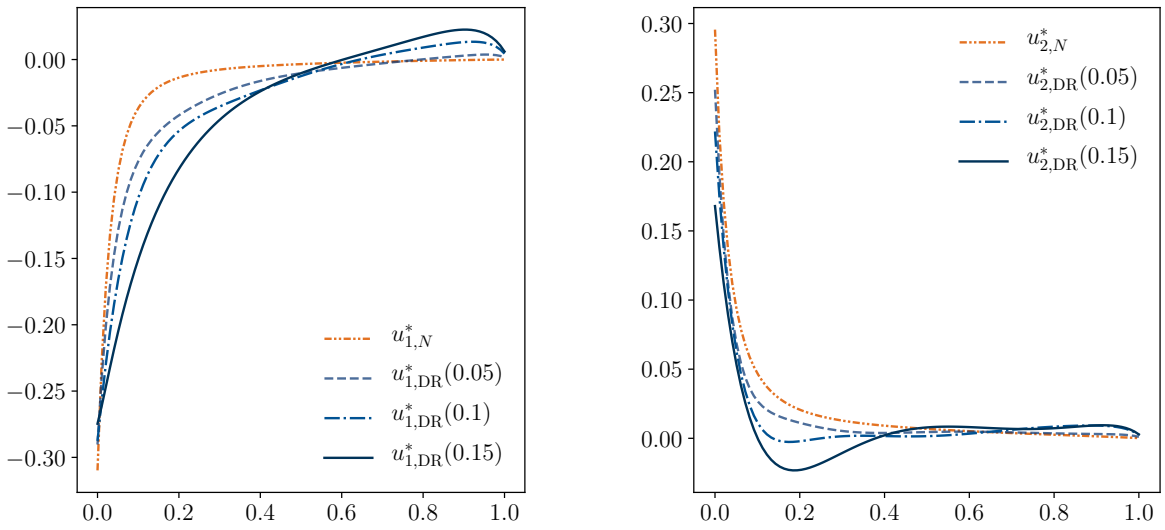


Figure 2.2: *Stationary control* $(u^*_{1,N}, u^*_{2,N})$ *of* (2.2.2) *and* $(u^*_{1,DR}(\Delta), u^*_{2,DR}(\Delta))$ *of* (2.1.5) *for* $\Delta = 0.05 = 10\sigma_1$, $\Delta = 0.1 = 10\sigma_1$, *and* $\Delta = 0.5 = 10\sigma_1$, *associated with the approximated DROP of the unsteady Burgers' equation* (2.7.10)*. The remaining data that defines the ambiguity set* $\mathcal{P}$ *(see* (2.1.2)*) is* $\sigma_0 = 0$, $\bar{\mu} = 0$, $\bar{\Sigma} = I$.

2.8 Conclusion and Discussion

We developed a sampling-free approximation scheme for moment-based distributionally robust PDE-constrained optimization problems. The definition of the moment-based ambiguity set provided in (2.1.2) is built on those used in [94, 300, 75]. Our approach incorporates second-order information from the reduced parameterized objective function (see (2.1.3)) about the nominal parameters into the problem formulation (2.1.5), and only requires one solution of the state equation per evaluation of the surrogate objective function.
In section 2.3, we provided conditions on the PDE solution and on the parameterized objective function that ensure the existence of optimal solutions of the DROP (2.1.1), and of the approximated and smoothed DROPs. To establish the existence of worst-case distributions, we used the concept of uniform integrability and we proved that the ambiguity set is weak-star sequentially compact.
In section 2.7, we applied our scheme to two nonlinear, nonconvex PDE-constrained problems: the optimal control of the steady and of the unsteady Burgers' equation. The optimal control of the unsteady Burgers' equation has applications in the field of PDE-constrained optimization as it allows the modeling of convection-diffusion phenomena [328, p. 69], [92, p. 203].
Future work includes the analysis of local convergence properties of the smoothing method, which may be built on the arguments used in section 1.5 combined with those in [193], and the application of our scheme to further control problems with a larger number of parameters.
A further task is the development of a scheme for data-driven DRO with PDEs using the Wasserstein distance. Such a scheme may be built on the use of second-order Taylor expansions or on utilizing Lagrangian relaxation as proposed by Sinha, Namkoong, and Duchi [299].

2.9 Supplementary Materials

2.9.1 Bounds on Moments of Sub-Gaussian Random Vectors

We prove upper bounds on the strong moments of sub-Gaussian random vectors.

Lemma 2.9.1 ([235, Lem. A.1]). *For all $P \in \mathcal{P}$, where the ambiguity set $\mathcal{P}$ is defined in* (2.1.2)*, and each $\gamma \geq 2$, we have $\mathbb{E}_P[\xi] \in \mathbb{R}^p$, and*

$$\mathbb{E}_P[\|\xi - \mathbb{E}_P[\xi]\|_2^\gamma] \leq 2(\gamma/\mathrm{e})^{\gamma/2}(I \bullet \sigma_1\bar{\Sigma})^{\gamma/2}, \tag{2.9.1}$$

$$\mathbb{E}_P[\|\xi\|_2^\gamma] \leq 2^\gamma(\gamma/\mathrm{e})^{\gamma/2}(I \bullet \sigma_1\bar{\Sigma})^{\gamma/2} + 2^{\gamma-1}(\|\bar{\Sigma}^{1/2}\|_2\Delta + \|\bar{\mu}\|_2)^\gamma. \tag{2.9.2}$$

Proof. Fix $P \in \mathcal{P}$. The definition of $\mathcal{P}$ provided in (2.1.2) ensures $\|\bar{\Sigma}^{-1/2}(\mathbb{E}_P[\xi] - \bar{\mu})\|_2 \leq \Delta$. We have $\mathbb{E}_P[\xi] \in \mathbb{R}^p$ since

$$\|\mathbb{E}_P[\xi]\|_2 \leq \|\bar{\Sigma}^{-1/2}\|_2\|\bar{\Sigma}^{-1/2}(\mathbb{E}_P[\xi] - \bar{\mu})\|_2 + \|\bar{\mu}\|_2 \leq \|\bar{\Sigma}^{-1/2}\|_2\Delta + \|\bar{\mu}\|_2 < \infty. \tag{2.9.3}$$

Combined with Lemma 1.8.2, we obtain the estimate (2.9.1).
The monotonicity and convexity of $\mathbb{R}_+ \ni z \mapsto z^\gamma$ yield, for each $\xi \in \mathbb{R}^p$,

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

Combined with (2.9.1) and (2.9.3), we obtain (2.9.2). □

Lemma 2.9.2 is used to verify Assumption 2.3.3 in section 2.7.

Lemma 2.9.2 ([235, Lem. A.3]). *Define $a : \mathbb{R}^p \to \mathbb{R}$ by $a(\xi) = \sum_{i=1}^p \alpha_i\xi_i^{r_i}$, and $b : \mathbb{R}^p \to \mathbb{R}$ by $b(\xi) = \exp\left(\sum_{i=1}^p \beta_i\xi_i\right)$, where α_i, $\beta_i \in \mathbb{R}$ and $r_i \geq 1$ are fixed. Then $|a|^r$, $|b|^s$, and $|a^t b^s|^r$ are uniformly integrable for all r, $t > 0$ and $s \in \mathbb{R}$.*

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

$$\begin{aligned} |a(\xi)|^{r(2\bar{r}/r)} = |a(\xi)|^{2\bar{r}} &\leq p^{2\bar{r}-1} \sum_{i=1}^{p} |\alpha_i|^{2\bar{r}} |\xi_i|^{2\bar{r}r_i} \leq p^{2\bar{r}-1} \sum_{i=1}^{p} |\alpha_i|^{2\bar{r}} \|\xi\|_2^{2\bar{r}r_i}, \\ |a(\xi)^t b(\xi)^s|^{2\bar{r}} &\leq |a(\xi)|^{2t\bar{r}} |b(\xi)|^{2s\bar{r}} \leq (1/2)|a(\xi)|^{4t\bar{r}} + (1/2)|b(\xi)|^{4s\bar{r}}. \end{aligned} \tag{2.9.4}$$

Combined with Lemmas 2.3.4 and 2.9.1, we find that $|a|^r$ is uniformly integrable for all $r > 0$. We have $|b(\xi)|^{2s} = \exp(\sum_{i=1}^{p} 2s\alpha_i \xi_i)$. We define $d \in \mathbb{R}^p$ by $d_i = 2s\alpha_i$ for $i = 1, \ldots, p$. Hence (2.3.2), Lemma 2.9.1, and (2.9.2) yield the uniform integrability of $|b|^s$ for all $s \in \mathbb{R}$. Moreover, the first two assertions, when combined with Lemma 2.3.4 and (2.9.4), imply the uniform integrability of $|a^t b^s|^r$ for each r, $t > 0$ and $s \in \mathbb{R}$. □

2.9.2 Weak-Weak Continuity of Solution Operators

We provide conditions that ensure the weak-weak continuity of the solution operator of a PDE. A similar result appears in the proof of [202, Thm. 12] by Kunisch and Walter [202].

Proposition 2.9.3 ([235, Lem. B.1])**.** *Let Y, U and Z be Banach spaces, let Y be reflexive, and let $U_{ad} \subset U$. For each $u \in U_{ad}$, let $S(u) \in Y$ be the unique solution to: Find $y \in Y$ with $e(y, u) = 0$, where $S : U_{ad} \to Y$ and $e : Y \times U \to Z$. Suppose that $\rho : \mathbb{R} \to \mathbb{R}$ is monotonically increasing with $\|S(u)\|_Y \leq \rho(\|u\|_U)$ for all $u \in U_{ad}$, and $X = \{ (y, u) \in Y \times U_{ad} : e(y, u) = 0 \}$ is weakly sequentially closed. Then S is weakly-weakly continuous.*

We apply the following known result to prove Proposition 2.9.3; cf. [196, p. 257], [277, pp. 236–238].

Lemma 2.9.4 ([235, Lem. B.2])**.** *Let X be a Banach space, and let $x \in X$. If each subsequence of $(x_k) \subset X$ has a further subsequence that converges weakly to x, then $x_k \rightharpoonup x$ as $k \to \infty$.*

Proof. For each fixed $f \in X^*$, each subsequence of $(\langle f, x_k \rangle_{X^*,X}) \subset \mathbb{R}$ has a further subsequence that converges to $\langle f, x \rangle_{X^*,X} \in \mathbb{R}$. Hence $\langle f, x_k \rangle_{X^*,X} \to \langle f, x \rangle_{X^*,X}$ as $k \to \infty$. □

Proof of Proposition 2.9.3. Fix $u \in U_{\mathrm{ad}}$ and $(u_k) \subset U_{\mathrm{ad}}$ with $u_k \rightharpoonup u$ as $k \to \infty$. Since ρ is increasing, the boundedness of (u_k) [46, Thm. 2.23] implies that of $(S(u_k))$. Let $(S(u_k))_K$ be a subsequence of $(S(u_k))$. Because $(S(u_k))$ is bounded and Y is reflexive, there exist $y \in Y$ and a further subsequence $(S(u_k))_{K'}$ of $(S(u_k))_K$ such that $S(u_k) \rightharpoonup y$ as $K' \ni k \to \infty$ [46, Thm. 2.28]. Since $(S(u_k), u_k)_{\mathbb{N}_0} \subset X$ and X is weakly sequentially closed, we deduce $(y, u) \in X$. Hence $y = S(u)$. Consequently, every subsequence of $(S(u_k))$ has a further subsequence converging to $S(u)$ weakly. Lemma 2.9.4 yields $S(u_k) \rightharpoonup S(u)$ as $k \to \infty$. □

2.9.3 Computation of Derivatives and Computational Complexity

Using the adjoint approach, we derive the derivatives required by our approximation scheme.

Adjoint Approach

We recap the adjoint approach (see, e.g., [151, sect. 1.6.4]), which we use to compute (first) derivatives of "reduced" functions.

Let Y, U and Z be Banach spaces, and let $\mathsf{F} : \mathsf{Y} \times \mathsf{U} \to \mathbb{R}$ and $\mathsf{E} : \mathsf{Y} \times \mathsf{U} \to \mathsf{Z}$ be continuously differentiable. We compute the derivative of $\widehat{\mathsf{F}} : \mathsf{U} \to \mathbb{R}$ defined by

$$\widehat{\mathsf{F}}(\mathsf{u}) = \mathsf{F}(\mathsf{S}(\mathsf{u}), \mathsf{u}).$$

For each $\mathsf{u} \in \mathsf{U}$, we assume that there exists a unique $\mathsf{y} = \mathsf{S}(\mathsf{u}) \in \mathsf{Y}$ such that

$$\mathsf{E}(\mathsf{y}, \mathsf{u}) = 0,$$

and require that $\mathsf{E}_\mathsf{y}(\mathsf{S}(\mathsf{u}), \mathsf{u})$ is boundedly invertible. The implicit function theorem implies that $\mathsf{S} : \mathsf{U} \to \mathsf{Y}$ is continuously differentiable [97, Thm. 10.2.1].
We define $\mathsf{L} : \mathsf{Y} \times \mathsf{U} \times \mathsf{Z}^* \to \mathbb{R}$ by

$$\mathsf{L}(\mathsf{y}, \mathsf{u}, \mathsf{z}) = \mathsf{F}(\mathsf{y}, \mathsf{u}) + \langle \mathsf{z}, \mathsf{E}(\mathsf{y}, \mathsf{u}) \rangle_{\mathsf{Z}^*, \mathsf{Z}}. \tag{2.9.5}$$

Using $\mathsf{E}(\mathsf{S}(\mathsf{u}), \mathsf{u}) = 0$, we have

$$\widehat{\mathsf{F}}(\mathsf{u}) = \mathsf{F}(\mathsf{y}(\mathsf{u}), \mathsf{u}) = \mathsf{L}(\mathsf{y}(\mathsf{u}), \mathsf{u}, \mathsf{z}) \quad \text{for all} \quad \mathsf{z} \in \mathsf{Z}.$$

We obtain

$$\widehat{\mathsf{F}}_\mathsf{u}(\mathsf{u}) = \mathsf{L}_\mathsf{u}(\mathsf{S}(\mathsf{u}), \mathsf{u}, \mathsf{z}(\mathsf{u})) = \mathsf{F}_\mathsf{u}(\mathsf{S}(\mathsf{u}), \mathsf{u}) + \mathsf{E}_\mathsf{u}(\mathsf{S}(\mathsf{u}), \mathsf{u})^* \mathsf{z}(\mathsf{u}), \tag{2.9.6}$$

where $\mathsf{z}(\mathsf{u}) \in \mathsf{Z}^*$ is the unique solution of the adjoint equation

$$\mathsf{L}_\mathsf{y}(\mathsf{y}(\mathsf{u}), \mathsf{u}, \mathsf{z}) = 0 \quad \Longleftrightarrow \quad \mathsf{E}_\mathsf{y}(\mathsf{y}(\mathsf{u}), \mathsf{u})^* \mathsf{z} = -\mathsf{F}_\mathsf{y}(\mathsf{y}(\mathsf{u}), \mathsf{u}), \tag{2.9.7}$$

see, e.g, [151, sects. 1.6.2 and 1.6.4].
In the subsequent sections, we exploit the symmetry of second-order partial derivatives of twice Fréchet differentiable mappings; see, e.g., [97, Thm. 8.12.2].

Computation of the Quadratic Model

We derive formulas for the first and second derivative of the function $\widehat{J}(u, \cdot)$ defined in (2.1.3) for fixed $u \in U$. These formulates can also be found in [151, sects. 1.6.2 and 1.6.5]. However, we use a different approach to compute the second derivative than that utilized in [151, sect. 1.6.5].
Throughout the current and next section, we require that the following conditions hold:

- The spaces U, Y, Z, and Ξ are Banach spaces.
- The mappings $J : U \times Y \times \Xi \to \mathbb{R}$ and $e : Y \times U \times \Xi \to Z$ are twice continuously differentiable.
- For each $(u, \xi) \in U \times \Xi$, $S(u, \xi) \in Y$ is the unique solution to: Find $y \in Y$ with $e(y, u, \xi) = 0$, where $S : U \times \Xi \to Y$.
- For each $(u, \xi) \in U \times \Xi$, the operator $e_y(S(u, \xi), u, \xi)$ is boundedly invertible.

(It would be sufficient to assume that U, Y, Z, and Ξ are open subsets of Banach spaces.)
We compute the first derivative of $\widehat{J}(u, \cdot)$. In order to be able to apply the adjoint approach described in section 2.9.3, we define $L : Y \times U \times \Xi \times Z^* \to \mathbb{R}$ by

$$L(y, u, \xi, z) = J(y, u, \xi) + \langle z, e(y, u, \xi) \rangle_{Z^*, Z}. \tag{2.9.8}$$

We identify $\mathsf{y} = y$, $\mathsf{u} = \xi$, $\mathsf{F} = J(\cdot, u, \cdot)$, $\mathsf{Y} = Y$, $\mathsf{U} = \Xi$, and keep $u \in U$ fixed. Applying the adjoint approach, and using (2.9.6) and (2.9.7), we obtain that $\widehat{J}_\xi(u, \xi) \in \Xi^*$ is given by

$$\widehat{J}_\xi(u, \xi) = J_\xi(S(u, \xi), u, \xi) + e_\xi(S(u, \xi), u, \xi)^* z_{\mathrm{foa}}(u, \xi), \tag{2.9.9}$$

where $z_{\mathrm{foa}}(u, \xi) \in Z^*$ is the solution of the *first-order adjoint equation*

$$e_y(S(u, \xi), u, \xi)^* z_{\mathrm{foa}}(u, \xi) = -J_y(S(u, \xi), u, \xi). \tag{2.9.10}$$

Hence, we can compute $\widehat{J}_\xi(u, \cdot)$ via the solution of the first-order adjoint equation (2.9.10).

Now, we compute the derivative of $\widehat{J}_\xi(u,\cdot)s_\xi$ for a fixed $s_\xi \in \Xi$. Let $y_{\text{fos}}(u,\xi;s_\xi) \in Y$ be the solution to the *first-order sensitivity equation*

$$e_y(S(u,\xi),u,\xi)y_{\text{fos}} = -e_\xi(S(u,\xi),u,\xi)s_\xi. \tag{2.9.11}$$

In order to apply the adjoint approach described in section 2.9.3, we define $\mathsf{y} = (y, y_{\text{fos}})$, $\mathsf{u} = \xi$, $\mathsf{z} = (z_1, z_2)$, $\mathsf{F}(\mathsf{y},\mathsf{u}) = J_y(y,u,\xi)y_{\text{fos}} + J_\xi(y,u,\xi)s_\xi$,

$$\mathsf{E}(\mathsf{y},\mathsf{u}) = \begin{bmatrix} e(y,u,\xi) \\ e_y(y,u,\xi)y_{\text{fos}} + e_\xi(y,u,\xi)s_\xi \end{bmatrix},$$

$\mathsf{Y} = Y \times Y$, $\mathsf{U} = \Xi$, $\mathsf{Z} = Z \times Z$, and fix $(s_\xi, u) \in \Xi \times U$. Here $\mathsf{y}(u) = (y(u,\xi), y_{\text{fos}}(u,\xi;s_\xi))$. We define $\mathsf{L} : \mathsf{Y} \times \mathsf{U} \times \mathsf{Z}^* \to \mathbb{R}$ by

$$\begin{aligned} \mathsf{L}(\mathsf{y},\mathsf{u},\mathsf{z}) = {} & J_y(y,u,\xi)y_{\text{fos}} + J_\xi(y,u,\xi)s_\xi + \langle z_1, e(y,u,\xi)\rangle_{Z^*,Z} \\ & + \langle z_2, e_y(y,u,\xi)y_{\text{fos}} + e_\xi(y,u,\xi)s_\xi\rangle_{Z^*,Z}. \end{aligned} \tag{2.9.12}$$

Combined with (2.9.11), we find that

$$\widehat{J}_\xi(u,\xi)s_\xi = J_y(S(u,\xi),u,\xi)y_{\text{fos}}(u,\xi;s_\xi) + J_\xi(S(u,\xi),u,\xi)s_\xi = \mathsf{L}(\mathsf{y}(\mathsf{u}),\mathsf{u},\mathsf{z}), \tag{2.9.13}$$

for all $\mathsf{z} = (z_1, z_2) \in \mathsf{Z}$. Using the adjoint approach, and (2.9.6) and (2.9.7), we obtain

$$\widehat{J}_{\xi\xi}(u,\xi)s_\xi = \mathsf{L}_\mathsf{u}(\mathsf{y}(\mathsf{u}),\mathsf{u},\mathsf{z}(\mathsf{u})), \tag{2.9.14}$$

where $\mathsf{z}(\mathsf{u}) \in \mathsf{Z}^*$ is the solution to

$$\mathsf{L}_\mathsf{y}(\mathsf{y}(\mathsf{u}),\mathsf{u},\mathsf{z}) = 0. \tag{2.9.15}$$

We compute the derivatives in (2.9.14) and (2.9.15). From (2.9.8), we obtain

$$\mathsf{L}(\mathsf{y},\mathsf{u},\mathsf{z}) = L_y(y,u,\xi,z_2)y_{\text{fos}} + L_\xi(y,u,\xi,z_2)s_\xi + \langle z_1, e(y,u,\xi)\rangle_{Z^*,Z}. \tag{2.9.16}$$

Recalling that $\mathsf{u} = \xi$ and $\mathsf{y} = (y, y_{\text{fos}})$, and using (2.9.8), we obtain

$$\begin{aligned} \mathsf{L}_\mathsf{u}(\mathsf{y}(\mathsf{u}),\mathsf{u},\mathsf{z}(\mathsf{u})) = {} & L_{y\xi}(S(u,\xi),u,\xi,z_{\text{foa}}(u,\xi))y_{\text{fos}}(u,\xi;s_\xi) \\ & + L_{\xi\xi}(S(u,\xi),u,\xi,z_{\text{foa}}(u,\xi))s_\xi + e_\xi(S(u,\xi),u,\xi)^* z_1, \\ \mathsf{L}_{\mathsf{y}_1}(\mathsf{y}(\mathsf{u}),\mathsf{u},\mathsf{z}(\mathsf{u})) = {} & L_{yy}(S(u,\xi),u,\xi,z_{\text{foa}}(u,\xi))y_{\text{fos}}(u,\xi;s_\xi) \\ & + L_{\xi y}(S(u,\xi),u,\xi,z_{\text{foa}}(u,\xi))s_\xi + e_y(S(u,\xi),u,\xi)^* z_1, \\ \mathsf{L}_{\mathsf{y}_2}(\mathsf{y}(\mathsf{u}),\mathsf{u},\mathsf{z}(\mathsf{u})) = {} & J_y(S(u,\xi),u,\xi) + e_y(S(u,\xi),u,\xi)^* z_2. \end{aligned} \tag{2.9.17}$$

We define $z_{\text{soa}}(u,\xi;s_\xi) \in Z^*$ as the solution to the *second-order adjoint equation*

$$\begin{aligned} e_y(S(u,\xi),u,\xi)^* z_{\text{soa}} = {} & -L_{yy}(S(u,\xi),u,\xi,z_{\text{foa}}(u,\xi))y_{\text{fos}}(u,\xi;s_\xi) \\ & - L_{\xi y}(S(u,\xi),u,\xi,z_{\text{foa}}(u,\xi))s_\xi. \end{aligned} \tag{2.9.18}$$

The equality (2.9.15) holds if $\mathsf{y}(\mathsf{u}) = (S(u,\xi), y_{\text{fos}}(u,\xi;s_\xi))$ and $\mathsf{z}(\mathsf{u}) = (z_{\text{soa}}(u,\xi), z_{\text{foa}}(u,\xi))$. Consequently $z_2 = z_{\text{foa}}(u,\xi)$. Our formulas agree with those obtained in [151, sect. 1.6.5].

Computation of the Derivative of the Quadratic Model

We derive formulas for the derivatives of $\widehat{J}(\cdot,\xi)$, $\widehat{J}_\xi(\cdot,\xi)s_\xi$ and of $\widehat{J}_{\xi\xi}(\cdot,\xi)[s_\xi,s_\xi]$ for fixed $(\xi,s_\xi)\in\Xi\times\Xi$. These formulates can also be found in [180, sect. 4.3] and [181, sect. 4.2]; however, our derivation differs.
We compute the derivative of $\widehat{J}(\cdot,\xi)$. The adjoint approach implies

$$\widehat{J}_u(u,\xi)=J_u(S(u,\xi),u,\xi)+e_u(S(u,\xi),u,\xi)^*z_{\mathrm{foa}}(u,\xi), \tag{2.9.19}$$

where $z_{\mathrm{foa}}(u,\xi)$ is the solution to (2.9.10).
We compute the derivative of $\widehat{J}_\xi(\cdot,\xi)s_\xi$. To apply the adjoint approach, we use the formula for $\widehat{J}_\xi(\cdot,\xi)s_\xi$ provided in (2.9.13). Here, however, we identify $\mathsf{u}=u$ and fix $(\xi,s_\xi)\in\Xi\times\Xi$. Applying the adjoint approach, and using (2.9.6) and (2.9.7), we find that

$$\widehat{J}_{\xi u}(u,\xi)s_\xi=\mathsf{L}_\mathsf{u}(\mathsf{y}(\mathsf{u}),\mathsf{u},\mathsf{z}(\mathsf{u})), \tag{2.9.20}$$

where $\mathsf{z}(\mathsf{u})\in\mathsf{Z}^*$ is the solution to

$$\mathsf{L}_\mathsf{y}(\mathsf{y}(\mathsf{u}),\mathsf{u},\mathsf{z})=0.$$

A formula for $\mathsf{L}_\mathsf{y}(\mathsf{y}(\mathsf{u}),\mathsf{u},\mathsf{z})$ is provided in (2.9.17). Using (2.9.16) and (2.9.20), we have

$$\begin{aligned}\widehat{J}_{\xi u}(u,\xi)s_\xi=&L_{yu}(S(u,\xi),u,\xi,z_{\mathrm{foa}}(u,\xi))y_{\mathrm{fos}}(u,\xi;s_\xi)\\&+L_{\xi u}(S(u,\xi),u,\xi,z_{\mathrm{foa}}(u,\xi))s_\xi\\&+e_u(y,u,\xi)^*z_{\mathrm{soa}}(u,\xi;s_\xi).\end{aligned}$$

We compute the derivative of $\widehat{J}_{\xi\xi}(\cdot,\xi)[s_\xi,s_\xi]$. Let $y_{\mathrm{sos}}(u,\xi;s_\xi)\in Y$ be the solution to the *second-order sensitivity equation*

$$\begin{aligned}e_y(S(u,\xi),u,\xi)y_{\mathrm{sos}}=&-e_{yy}(S(u,\xi),u,\xi)[y_{\mathrm{fos}}(u,\xi;s_\xi),y_{\mathrm{fos}}(u,\xi;s_\xi)]\\&-2e_{\xi y}(S(u,\xi),u,\xi)[s_\xi,y_{\mathrm{fos}}(u,\xi;s_\xi)]\\&-e_{\xi\xi}(S(u,\xi),u,\xi)[s_\xi,s_\xi].\end{aligned} \tag{2.9.21}$$

We identify $\mathsf{y}=(y,y_{\mathrm{fos}},y_{\mathrm{sos}})$, $\mathsf{y}(\mathsf{u})=(S(u,\xi),y_{\mathrm{fos}}(u,\xi;s_\xi),y_{\mathrm{sos}}(u,\xi;s_\xi))$, $\mathsf{u}=u$, $\mathsf{z}=(z_1,z_2,z_3)\in\mathsf{Z}^*=Z^*\times Z^*\times Z^*$, $\mathsf{F}(\mathsf{y},\mathsf{u})=J_y(y,u,\xi)y_{\mathrm{sos}}$,

$$\mathsf{E}(\mathsf{y},\mathsf{u})=\begin{bmatrix}e(y,u,\xi)\\ e_y(y,u,\xi)y_{\mathrm{fos}}+e_\xi(y,u,\xi)s_\xi\\ e_y(y,u,\xi)y_{\mathrm{sos}}+e_{yy}(y,u,\xi)[y_{\mathrm{fos}},y_{\mathrm{fos}}]+2e_{\xi y}(y,u,\xi)[s_\xi,y_{\mathrm{fos}}]+e_{\xi\xi}(y,u,\xi)[s_\xi,s_\xi]\end{bmatrix},$$

and fix $(\xi,s_\xi)\in\Xi\times\Xi$. We define $\mathsf{M}:\mathsf{Y}\times\mathsf{U}\times\mathsf{Z}^*\to\mathbb{R}$ by

$$\begin{aligned}\mathsf{M}(\mathsf{y},\mathsf{u},\mathsf{z})=&J_y(y,u,\xi)y_{\mathrm{sos}}\\&+J_{yy}(y,u,\xi)[y_{\mathrm{fos}},y_{\mathrm{fos}}]+2J_{y\xi}(y,u,\xi)[y_{\mathrm{fos}},s_\xi]+J_{\xi\xi}(y,u,\xi)[s_\xi,s_\xi]\\&+\langle z_1,e(y,u,\xi)\rangle_{Z^*,Z}\\&+\langle z_2,e_y(y,u,\xi)y_{\mathrm{fos}}+e_\xi(y,u,\xi)s_\xi\rangle_{Z^*,Z}\\&+\langle z_3,e_y(y,u,\xi)y_{\mathrm{sos}}+e_{yy}(y,u,\xi)[y_{\mathrm{fos}},y_{\mathrm{fos}}]\rangle_{Z^*,Z}\\&+\langle z_3,2e_{\xi y}(y,u,\xi)[s_\xi,y_{\mathrm{fos}}]+e_{\xi\xi}(y,u,\xi)[s_\xi,s_\xi]\rangle_{Z^*,Z}.\end{aligned} \tag{2.9.22}$$

Combining (2.9.11), (2.9.21), and (2.9.22), we find that

$$\begin{aligned}\widehat{J}_{\xi\xi}(u,\xi)[s_\xi,s_\xi]=&J_y(S(u,\xi),u,\xi)y_{\mathrm{sos}}(u,\xi;s_\xi)+J_{yy}(S(u,\xi),u,\xi)[y_{\mathrm{fos}}(u,\xi;s_\xi),y_{\mathrm{fos}}(u,\xi;s_\xi)]\\&+2J_{y\xi}(S(u,\xi),u,\xi)[y_{\mathrm{fos}}(u,\xi;s_\xi)s_\xi,s_\xi]+J_{\xi\xi}(S(u,\xi),u,\xi)[s_\xi,s_\xi]\\=&\mathsf{M}(\mathsf{y}(\mathsf{u}),\mathsf{u},\mathsf{z}),\end{aligned}$$

for all $\mathsf{z} \in \mathsf{Z}$. Applying the adjoint approach, and using (2.9.6) and (2.9.7), we obtain

$$\widehat{J}_{\xi\xi u}(u,\xi)[s_\xi, s_\xi] = \mathsf{M}_\mathsf{u}(\mathsf{y}(\mathsf{u}), \mathsf{u}, \mathsf{z}(\mathsf{u})), \tag{2.9.23}$$

where $\mathsf{z}(\mathsf{u}) \in \mathsf{Z}^*$ is the solution to

$$\mathsf{M}_\mathsf{y}(\mathsf{y}(\mathsf{u}), \mathsf{u}, \mathsf{z}) = 0. \tag{2.9.24}$$

Using (2.9.8) and (2.9.21), we have

$$\begin{aligned}\mathsf{M}(\mathsf{y},\mathsf{u},\mathsf{z}) = {} & L_y(y,u,\xi,z_3)y_{\mathrm{sos}} + L_{yy}(y,u,\xi,z_3)[y_{\mathrm{fos}}, y_{\mathrm{fos}}] \\ & + 2L_{y\xi}(y,u,\xi,z_3)[y_{\mathrm{fos}}, s_\xi] + L_{\xi\xi}(y,u,\xi,z_3)[s_\xi, s_\xi] \\ & + \langle z_1, e(y,u,\xi)\rangle_{Z^*,Z} + \langle z_2, e_y(y,u,\xi)y_{\mathrm{fos}} + e_\xi(y,u,\xi)s_\xi\rangle_{Z^*,Z}.\end{aligned} \tag{2.9.25}$$

From (2.9.22) and the chain rule, we deduce

$$\begin{aligned}\mathsf{M}_{\mathsf{y}_1}(\mathsf{y},\mathsf{u},\mathsf{z}) = {} & L_{yy}(y,u,\xi,z_3)y_{\mathrm{sos}} + L_{yyy}(y,u,\xi,z_3)[y_{\mathrm{fos}}, y_{\mathrm{fos}}, \cdot] \\ & + 2L_{y\xi y}(y,u,\xi,z_3)[y_{\mathrm{fos}}, s_\xi, \cdot] + L_{\xi\xi y}(y,u,\xi,z_3)[s_\xi, s_\xi, \cdot] \\ & + e_y(y,u,\xi)^* z_1 \\ & + [e_{yy}(y,u,\xi)y_{\mathrm{fos}}]^* z_2 + [e_{\xi y}(y,u,\xi)s_\xi]^* z_2, \\ \mathsf{M}_{\mathsf{y}_2}(\mathsf{y},\mathsf{u},\mathsf{z}) = {} & 2L_{yy}(y,u,\xi)y_{\mathrm{fos}} + 2L_{y\xi}(y,u,\xi)s_\xi + e_y(y,u,\xi)^* z_2, \\ \mathsf{M}_{\mathsf{y}_3}(\mathsf{y},\mathsf{u},\mathsf{z}) = {} & L_y(y,u,\xi,z_3).\end{aligned}$$

Using (2.9.10), we get $z_3(u,\xi) = z_{\mathrm{foa}}(u,\xi)$, and (2.9.18) yields $z_2 = z_{\mathrm{soa}}(u,\xi;s_\xi)$. Now, we define $z_{\mathrm{toa}}(u,\xi;s_\xi) \in Z^*$ as the solution of the *third-order adjoint equation*

$$\begin{aligned}e_y(S(u,\xi),u,\xi)^* z_{\mathrm{toa}} = {} & -L_{yy}(S(u,\xi),u,\xi,z_{\mathrm{foa}}(u,\xi))y_{\mathrm{sos}}(u,\xi;s_\xi) \\ & - L_{yyy}(S(u,\xi),u,\xi,z_{\mathrm{foa}}(u,\xi))[y_{\mathrm{fos}}(u,\xi;s_\xi), y_{\mathrm{fos}}(u,\xi;s_\xi)] \\ & - 2L_{y\xi y}(S(u,\xi),u,\xi,z_{\mathrm{foa}}(u,\xi))[y_{\mathrm{fos}}(u,\xi;s_\xi), s_\xi, \cdot] \\ & - L_{\xi\xi y}(S(u,\xi),u,\xi,z_{\mathrm{foa}}(u,\xi))[s_\xi, s_\xi, \cdot] \\ & - [e_{yy}(S(u,\xi),u,\xi)y_{\mathrm{fos}}(u,\xi;s_\xi)]^* z_{\mathrm{soa}}(u,\xi;s_\xi) \\ & - [e_{\xi y}(S(u,\xi),u,\xi)s_\xi]^* z_{\mathrm{soa}}(u,\xi;s_\xi).\end{aligned} \tag{2.9.26}$$

Using (2.9.22) and (2.9.23), we find that

$$\begin{aligned}\widehat{J}_{\xi\xi u}(u,\xi)[s_\xi, s_\xi] = {} & L_{yu}(S(u,\xi),u,\xi,z_{\mathrm{foa}}(u,\xi))y_{\mathrm{sos}}(u,\xi;s_\xi) \\ & + L_{yyu}(S(u,\xi),u,\xi,z_{\mathrm{foa}}(u,\xi))[y_{\mathrm{fos}}(u,\xi;s_\xi), y_{\mathrm{fos}}(u,\xi;s_\xi), \cdot] \\ & + 2L_{y\xi u}(S(u,\xi),u,\xi,z_{\mathrm{foa}}(u,\xi))[y_{\mathrm{sos}}(u,\xi;s_\xi), s_\xi, \cdot] \\ & + L_{\xi\xi u}(S(u,\xi),u,\xi,z_{\mathrm{foa}}(u,\xi))[s_\xi, s_\xi, \cdot] \\ & + e_u(S(u,\xi),u,\xi)^* z_{\mathrm{toa}}(u,\xi) \\ & + [e_{yu}(S(u,\xi),u,\xi)y_{\mathrm{fos}}(u,\xi;s_\xi)]^* z_{\mathrm{soa}}(u,\xi;s_\xi) \\ & + [e_{\xi u}(S(u,\xi),u,\xi)s_\xi]^* z_{\mathrm{soa}}(u,\xi;s_\xi).\end{aligned} \tag{2.9.27}$$

Computation of Derivatives and Computational Complexity

We derive the number of solutions of linear and of nonlinear PDEs required to evaluate the smoothing function $\widetilde{F}$ defined in (2.2.1) and its derivative. We fix $t = (\tau, \nu, \eta) \in \mathbb{R}^3_{++}$.
To compute $\widetilde{F}(\bar{u};t)$, we evaluate $\widehat{J}(\bar{u},\bar{\mu})$, $\nabla_\xi \widehat{J}(\bar{u},\bar{\mu})$ and $\nabla_{\xi\xi}\widehat{J}(\bar{u},\bar{\mu})$. We compute $\widehat{J}(\bar{u},\bar{\mu})$ using the solution of the state equation. The gradient $\nabla_\xi \widehat{J}(\bar{u},\bar{\mu})$ can be computed via the adjoint

approach which requires the solution of the first-order adjoint equation (2.9.10). We compute the Hessian $\nabla_{\xi\xi}\widehat{J}(\bar{u},\bar{\mu})$ via the adjoint approach which requires the solution of p first-order sensitivity equations (see (2.9.11)) and of p second-order adjoint equations (see (2.9.18)).
To evaluate the derivative of $\widetilde{F}(\cdot;t)$ at $\bar{u}$, we use (2.6.4), (2.6.5) and (2.6.7), and compute $\mathrm{D}_u\widehat{J}(\bar{u},\bar{\mu})$, $\mathrm{D}_u(\nabla_\xi\widehat{J}(\bar{u},\bar{\mu})^T s_\xi)$, $\mathrm{D}_u(s_\xi^T\nabla_{\xi\xi}\widehat{J}(\bar{u},\bar{\mu})s_\xi)$, and $\mathrm{D}_u(q_i(\bar{u})^T\nabla_{\xi\xi}\widehat{J}(\bar{u},\bar{\mu})q_i(\bar{u}))$ for a certain $s_\xi \in \mathbb{R}^p$ and $i = 1,\dots,p$. Using (2.9.19), we can evaluate $\mathrm{D}_u\widehat{J}(\bar{u},\bar{\mu})$ without further costs. For $i = 1,\dots,p$, we compute $\mathrm{D}_u(q_i(\bar{u})^T\nabla_{\xi\xi}\widehat{J}(\bar{u},\bar{\mu})q_i(\bar{u}))$ using (2.9.27), requiring the solution of the second-order sensitivity equation (2.9.21), the second-order adjoint equation and the third-order adjoint equation (2.9.26). These PDE solutions allow us to compute $\mathrm{D}_u(\nabla\widehat{J}(\bar{u},\bar{\mu})^T s_\xi)$ using (2.9.23), and $\mathrm{D}_u(s_\xi^T\nabla_{\xi\xi}\widehat{J}(\bar{u},\bar{\mu})s_\xi)$ without additional costs.
To summarize, the evaluation of $F(\bar{u};t)$ involves the solution of the state equation, and $2p+1$ solutions of linear equations, and that of $\mathrm{D}_u F(\bar{u};t)$ requires, in addition, $2p$ solutions of linear equations. In order to compute the Riesz representation of $\mathrm{D}_u F(\bar{u};t)$, we need to solve one more linear PDE.

3 Sample Average Approximation for Stochastic Convex Optimal Control Problems: Non-Asymptotic Sample Size Estimates

We apply the sample average approximation (SAA) approach to stochastic, convex optimal control problems posed in Hilbert spaces. For strongly convex problems, we establish non-asymptotic, exponential bounds on the tail probabilities of the optimal controls. Non-asymptotic confidence intervals for the optimal value of the SAA problem are established without strong convexity assumptions. We demonstrate that our assumptions hold true for many PDE-constrained optimization problems under uncertainty with affine-linear elliptic and parabolic PDEs from the literature. A further focus is on the comparison of our bounds with those obtained from (robust) stochastic approximation. Furthermore, we apply the finite element discretization to a strongly convex optimal control problem with a random elliptic PDE, and derive confidence regions for the optimal control.

3.1 Introduction

The sample average approximation (SAA) method is an approach for approximating stochastic programs [287, 156, 347, 177, 179, 190]. The approach uses samples of the random vector to approximate the stochastic program's objective function using the sample average, thereby defining the sample average function [291, p. 355]. This function is the objective function of the SAA problem. The SAA problem's optimal value and its optimal solutions provide approximations to those of the stochastic program. We analyze the accuracy of these approximations for a class of risk-neutral control problems posed in Hilbert spaces.
We consider the risk-neutral control problem

$$\min_{u\in U_{\mathrm{ad}}} \{\, f(u) = \mathbb{E}[\widehat{J}(u,\xi)] + \Psi(u) \,\}, \tag{3.1.1}$$

where U_{ad} is a convex, closed and nonempty subset of the separable Hilbert space U, and $\widehat{J} : U \times \Xi \to \mathbb{R}$ is a convex Carathéodory function. Moreover, ξ is a random vector and its probability distribution is supported on some set Ξ. The penalty function $\Psi : U_{\mathrm{ad}} \to \mathbb{R} \cup \{\infty\}$ is convex and lower-semicontinuous.
The SAA problem corresponding to (3.1.1) is

$$\min_{u\in U_{\mathrm{ad}}} \big\{\, f_N(u) = \mathrm{E}^N[\widehat{J}(u,\xi)] + \Psi(u) \,\big\}, \tag{3.1.2}$$

where $\mathrm{E}^N[\widehat{J}(u,\xi)] = (1/N)\sum_{i=1}^N \widehat{J}(u,\xi^i)$, and $\xi^1, \xi^2, \ldots$ are independent samples and each ξ^i has the same probability distribution as ξ. For the analysis of the SAA approach in section 3.2, we view ξ^i as random vectors defined on a common probability space.
A central question related to the SAA approach is: how many samples must be used in order for an (approximate) optimal solution of the SAA problem to be an accurate solution to the true

counterpart? To address this question, we first need to choose a suitable measure to quantify accuracy. When approximating PDE-constrained optimization problems using finite elements—an approach, which also results in perturbed optimization problems—the main focus is on deriving bounds for the error between the optimal control of the finite element approximation and that of the unperturbed control problem [326, sect. 3], [315, sect. 2], [151, sect. 3.2.6], [65, sect. 1]. It will turn out that estimating the tail probabilities

$$\text{Prob}(\|u_N^* - u^*\|_U \geq \varepsilon) \quad \text{for} \quad \varepsilon > 0, \tag{3.1.3}$$

is the canonical extension for analyzing the accuracy of optimal solutions to the SAA problem. Here, u^* is a optimal control of (3.1.1) and u_N^* is a minimizer of the corresponding SAA problem (3.1.2).

When a bound on the tail probabilities (3.1.3) is available as a function of N, we can determine the number of samples N such that $\text{Prob}(\|u_N^* - u^*\|_U \geq \varepsilon) \leq \delta$. Here, $1 - \delta \in (0, 1)$ is the *reliability*, and $\varepsilon > 0$ the *accuracy*. One approach to obtaining bounds on (3.1.3) exploits Tschebyshev's inequality. For example, suppose that $\mathbb{E}[\|u_N^* - u^*\|_U^2] \leq C/N$ for some $C > 0$; then Tschebyshev's inequality yields $\text{Prob}(\|u_N^* - u^*\|_U \geq \varepsilon) \leq C/(N\varepsilon^2)$. If $N \geq C/(\varepsilon^2\delta)$, we obtain $\text{Prob}(\|u_N^* - u^*\|_U \geq \varepsilon) \leq \delta$. However, for $0 < \delta \ll 1$, the non-asymptotic sample size estimate $N \geq C/(\varepsilon^2\delta)$ would yield an infeasibly large number of samples.

Our main contribution is the derivation of an exponential bound on $\text{Prob}(\|u_N^* - u^*\|_U \geq \varepsilon)$, that is, $\text{Prob}(\|u_N^* - u^*\|_U \geq \varepsilon) \leq \exp(-N\varepsilon^2/c)$ for all $\varepsilon > 0$, and some problem-dependent constant $c > 0$. When $N \geq (c/\varepsilon^2)\ln(1/\delta)$, the tail bound ensures $\text{Prob}(\|u_N^* - u^*\|_U \geq \varepsilon) \leq \delta$. In contrast to the above sample size estimate obtained via the direct application of Tschebyshev's inequality, the non-asymptotic sample size estimate $N \geq \ln(1/\delta)(c/\varepsilon^2)$ depends only logarithmically on $1/\delta$.

We obtain the exponential bound on the tail probabilities (3.1.3) using a standard assumption from the literature on stochastic programming. We assume that $\widehat{J}(\cdot, \xi)$ for all $\xi \in \Xi$ and $\mathbb{E}[\widehat{J}(\cdot, \xi)]$ are Gâteaux differentiable, and that there exists $\tau > 0$ with

$$\mathbb{E}[\exp(\tau^{-2}\|\nabla_u \widehat{J}(u^*, \xi) - \nabla_u \mathbb{E}[\widehat{J}(u^*, \xi)]\|_U^2)] \leq \text{e}. \tag{3.1.4}$$

This condition and its variants are used in [207, eq. (4.1.15)], [99, p. 679], [243, eq. (2.50)], [138, pp. 1035–1036], and [294, eq. (5.347)], for example. We could impose an upper bound $c > 1$ on the left-hand side in (3.1.4) other than $c = \text{e}$, but it would only modify the constant $\tau > 0$.

In addition to (3.1.4), we suppose that $\widehat{J}(\cdot, \xi)$ is α-strongly convex for all $\xi \in \Xi$ with $\alpha > 0$. Under these assumptions, we establish the exponential tail bound

$$\text{Prob}(\|u^* - u_N^*\|_U \geq \varepsilon) \leq 2\exp(-\tau^{-2}N\varepsilon^2\alpha^2/3) \quad \text{for all} \quad \varepsilon > 0. \tag{3.1.5}$$

The "light-tail" condition (3.1.4) expresses sub-Gaussianity of $\nabla_u \widehat{J}(u^*, \xi) - \nabla_u \mathbb{E}[\widehat{J}(u^*, \xi)]$. It is fulfilled if $\|\nabla_u \widehat{J}(u^*, \xi) - \nabla_u \mathbb{E}[\widehat{J}(u^*, \xi)]\|_U$ is essentially bounded, for example. This property is satisfied for the control problems considered in [230, sect. 3.1.2], [227, sect. 2], [184, sect. 3.4.1], [139, sect. 2.2], [131, sect. 4], [125, sect. 2.2], and [308, sect. 2.2]. These control problems also satisfy the requirement of $\widehat{J}(\cdot, \xi)$ being α-strongly convex for all $\xi \in \Xi$. A discussion on the strong convexity assumption for stochastic, linear-quadratic control problems is provided in section 3.2.5. We also demonstrate that the tail bound (3.1.5) is optimal up to problem-independent, moderate constants, under the stated assumptions.

Besides deriving an exponential bound on the tail probabilities (3.1.3), we quantify the errors between the optimal value of the SAA problem (3.1.2) and that of the true counterpart (3.1.1) without the strong convexity assumption. More precisely, we derive exponential tail bounds on $f_N(u_N^*) - f(u^*)$. For a linear-quadratic control problem with a nonsmooth regularizer, we discretize the SAA problem using finite elements and derive reliable error bounds on the optimal

controls. Furthermore, we analyze the expected value of the SAA optimal value of a risk-averse optimization problem using the conditional value-at-risk/superquantile.
The derivation of the exponential bound (3.1.5) is rather simple and consists of three steps. First, we establish an exponential tail bound for the sample mean in Hilbert spaces. To derive this estimate, we combine the Chernoff-type approach with the (exponential) moment inequality established by Pinelis and Sakhanenko [259, Thm. 3]. Second, we derive an almost sure bound on $\|u_N^* - u^*\|_U$. Finally, the exponential tail bound is applied to the right-hand side of this bound.
Our approach for establishing exponential bounds on $f_N(u_N^*) - f(u^*)$ is built on the analysis developed by Guigues, Juditsky, and Nemirovski [138, Prop. 1].

Related Work

Non-Asymptotic Analysis of the SAA Scheme for Finite-Dimensional Problems. The complexity of finite-dimensional optimization problems has been analyzed by Shapiro [291, 292] and Shapiro and Nemirovski [296]. To outline one of their results, let us consider the stochastic problem and its SAA problem

$$\min_{x \in X_{\mathrm{ad}}} \mathbb{E}[F(x,\xi)] \quad \text{and} \quad \min_{x \in X_{\mathrm{ad}}} \mathrm{E}^N[F(x,\xi)], \tag{3.1.6}$$

where $F : X_{\mathrm{ad}} \times \Xi \to \mathbb{R}$, $\mathrm{E}^N[F(x,\xi)] = (1/N)\sum_{i=1}^N F(x,\xi^i)$, and ξ^i are independent ($i = 1,\ldots,N$) and each ξ^i has the same probability distribution as $\xi \in \Xi$ with $\Xi \subset \mathbb{R}^m$. We briefly outline the main assumptions made in [292, pp. 186 and 189] and [296, pp. 116–121]:

- The set $X_{\mathrm{ad}} \subset \mathbb{R}^n$ is nonempty and closed, and has finite diameter $R_{\mathrm{ad}} = \sup_{x_1, x_2 \in X_{\mathrm{ad}}} \|x_1 - x_2\| < \infty$. The function $\mathbb{E}[F(\cdot,\xi)]$ is well-defined and finite-valued on X_{ad}.
- For all x, $y \in X_{\mathrm{ad}}$, $F(x,\xi) - F(y,\xi) - \mathbb{E}[F(x,\xi) - F(y,\xi)]$ is sub-Gaussian with parameter $\sigma > 0$. (Sub-Gaussian random variables are defined on p. ix.)
- There exists a random variable $L : \Xi \to \mathbb{R}_+$ such that $\mathbb{E}[\exp(tL(\xi))] < \infty$ for all t in a neighborhood of 0, and for each $\xi \in \Xi$, $F(\cdot,\xi)$ is Lipschitz continuous w.r.t. the $\|\cdot\|$-norm with Lipschitz constant $L(\xi)$.

If these conditions are fulfilled, $0 \le \rho < \varepsilon$, $\delta \in (0,1)$, $\widetilde{L} > \mathbb{E}[L(\xi)]$, and

$$N \ge \max\left\{ \frac{8\sigma^2}{(\varepsilon-\rho)^2}\Big[n \ln\Big(\frac{c_1 \mathbb{E}[L(\xi)] R_{\mathrm{ad}}}{\varepsilon - \rho}\Big) + \ln\Big(\frac{1}{\delta}\Big)\Big],\ \frac{1}{c_2(\widetilde{L})} \ln\Big(\frac{2}{\delta}\Big) \right\}, \tag{3.1.7}$$

then a ρ-optimal solution of the SAA problem in (3.1.6) is an ε-optimal solution of the stochastic problem in (3.1.6) with a probability of at least $1 - \delta$ [291, sect. 3.2], [296, Thm. 1], [292, Thm. 1]. Here, $c_1 > 0$ is a problem-independent constant and $c_2(\widetilde{L}) > 0$ depends on $\widetilde{L}$ [292, eq. (3.6)]. The estimate (3.1.7) depends explicitly on the problem's dimension, and therefore cannot be applied to infinite-dimensional stochastic programs. Shapiro [292, Ex. 1], and Guigues, Juditsky, and Nemirovski [138, Prop. 2] provide examples of convex stochastic problems, which highlight the fact that the estimate (3.1.7) optimally depends on the problem's dimension n.
Even though the sample size estimate (3.1.7) depends on the problem's dimension, the proof technique developed in [291, 296] may be extended to stochastic programs with totally bounded feasible sets posed in infinite-dimensional spaces. The proof technique exploits the fact that the feasible set X_{ad} is a subset of a compact subset of $\mathbb{R}^n$ for which the ν-covering number w.r.t. the $\|\cdot\|$-norm is proportional to $(R_{\mathrm{ad}}/\nu)^n$ [296, p. 119].
Some function classes have finite covering numbers, including certain classes of the functions of bounded variation w.r.t. the L^1-norm [17, sect. 2], and certain collections of $\mathbb{R}^n$-valued upper-semicontinuous functions w.r.t. the Attouch–Wets distance [278, sect. 4]. Many subsets in

infinite-dimensional, complete spaces are noncompact and, hence, not totally bounded [196, p. 412], such as closed unit balls [196, Thm. 2.5-5]. Even further, the box-constrained set $\{ u \in L^2((0,1)) : -1 \leq u \leq 1 \}$—a common type of a feasible set considered in the literature on optimal control with PDEs [151, p. 71], [303, p. 160]—is noncompact, as it contains the non-convergent sequence $(\sin(k\pi\cdot))$.
Birman and Solomjak [40, Thm. 5.2], [41, Thms. 1.7 and 2.24] analyze the covering numbers of compact embeddings, from Sobolev spaces to Lebesgue spaces and to the space of continuous functions. As a special case, the ν-covering number of the closed unit ball of $H^1([-1,1]^d)$ w.r.t. the $L^2([-1,1]^d)$-norm is proportional to $(1/\nu)^d$. We outline in section 3.5 how this deep statement may be used to analyze the complexity of certain PDE-constrained optimization problems under uncertainty posed in the Lebesgue space $L^2(\mathcal{D})$, where $\mathcal{D} \subset \mathbb{R}^d$ is a bounded domain. For this discussion, we use the fact that the optimal controls of many (deterministic) PDE-constrained optimization problems posed in $L^2(\mathcal{D})$ are actually contained in $H^1(\mathcal{D})$, under suitable assumptions about the problem's data [335, p. 870], [65, Lem. 1.1], [316, Thms. 2.37 and 2.38].

Further Performances Guarantees for SAA Estimators. In statistics, the consistency of an estimator refers to its asymptotic, almost sure convergence to its true counterpart. For finite-dimensional problems, we refer the reader to Shapiro [291, sect. 2.1] for the consistency analysis of the SAA problem's optimal value and its solution (set); see also [294, sect. 5.1.1]. See Shapiro [289] for the asymptotic analysis of the SAA problem's optimal value.
The consistency for the SAA approach, applied to stochastic programs posed in complete, separable, metric spaces, is proven by Artstein and Wets [10, Thm. 2.3] using the notion of epiconvergence. Dong and Wets [98, Thm. 5.3] establish the consistency of the SAA method in Hilbert spaces using Mosco-epiconvergence, a stronger notion than epiconvergence. Phelps, Royset, and Gong [254] apply the SAA scheme to the optimal control of ordinary differential equations with random inputs, and analyze the consistency of the SAA optimal value using [10, Thm. 2.3]. Non-asymptotic error bounds using the (large deviations) rate function are provided in [170, Thm. 4.6] for stochastic problems posed in Banach spaces, but the assumptions of this result may be difficult to verify for PDE-constrained control problems under uncertainty.
Hoffhues, Römisch, and Surowiec [153] provide qualitative and quantitative stability results for the optimal value and for the optimal solutions of stochastic, linear-quadratic optimization problems posed in Hilbert spaces w.r.t. the Fortet–Mourier and Wasserstein distances.

Optimization Methods for Stochastic Programs. The SAA approach yields an approximated optimization problem, which can be interpreted as a stochastic program [294, pp. 163–164], and requires a numerical scheme for its solution. Different approaches for solving risk-neutral and risk-averse optimization problems have been proposed in the literature, such as stochastic approximation [128, 246, 243, 208], progressive hedging [270, 275], primal-dual subgradient methods [168, 248], and primal-dual multiplier-type algorithms [194, 282]. We highlight the work by Nemirovski and Yudin [246, Chap. 5] on stochastic approximation in infinite-dimensional spaces whose duals are so-called regular Banach spaces. Hilbert spaces are the simplest examples of such spaces [246, sect. 3.2]. Lan [207, Chap. 6] investigates stochastic approximation schemes applied to finite-dimensional, nonconvex, stochastic programs. Geiersbach and Scarinci [129] have developed stochastic gradient methods for nonconvex optimization problems posed in Hilbert spaces. See [127] for stochastic approximation applied to shape optimization under uncertainty.

Optimization Methods for PDE-Constrained Optimization under Uncertainty. In the literature on PDE-constrained optimization, several schemes have been developed to approximate, discretize and solve control problems with random inputs. We refer the reader to Kouri

and Shapiro [190, sect. 5] for a recent survey on methods for expectation-based optimization with a focus on control problems in Hilbert spaces.
Garreis and Ulbrich [123] view the parameterized state variable as a tensor space element and discretize it using finite elements and polynomials, and approximate the discretized state using low-rank tensor formats. Garreis [121] has developed error estimates, which quantify approximations in the objective function and gradient computations, and has created a trust-region algorithm to adaptively solve risk-neutral control problems. Tensor-based methods have also been developed in [29, 124].
Stochastic collocation and sparse grids provide a different discretization approach than the SAA method [51, 314, 184, 189, 188, 185]. Adaptive trust-region methods using inexact objective and gradient evaluations have been developed in [184, 189, 188, 185]. Kouri [184] has proven control error estimates for different risk-measures, such as the expectation [184, Thm. 3.4.1 and Cor. 3.4.2], the superquantile/conditional value-at-risk [184, Thm. 3.4.5], and the mean-plus-semideviations [184, Thm. 3.4.4]. A globally convergent optimization method using adaptive model reduction and sparse grids can be found in Zahr, Carlberg, and Kouri [357]. Kouri [187] has proposed using quadrature to approximate risk measures and has provided asymptotic consistency results. A quasi-Monte Carlo method for a risk-neutral, elliptic control problem has been developed in [140]. The authors have provided an error analysis and convergence rates w.r.t. truncation, finite element, and quadrature errors.
Stochastic gradient methods for PDE-constrained optimization under uncertainty can be found in [131, 128, 227, 228, 229]. Geiersbach and Wollner [131] have designed a stochastic gradient method with adaptive mesh refinement. Martin [228] and Martin, Krumscheid, and Nobile [227] provide error estimates in a mean-square sense for various inaccuracies, such as stochastic errors [227, Thm. 10] and finite element errors [227, Thm. 7]. A multilevel stochastic gradient method is developed in [229]. In the context of parameter estimation and inversion with PDEs, stochastic approximation and SAA are compared in [141, 219].

Large Deviations, Moment Inequalities, and Exponential Tail Bounds. We prove a large deviation result using the exponential moment inequality established by Pinelis and Sakhanenko [259, Thm. 3]. A large deviation result for a certain class of Banach spaces has been announced by Nemirovski [241, Chap. 3] and proven for (finite-dimensional) Banach spaces by Juditsky and Nemirovski [166] using an "optimization-based" proof. Large deviations results for certain Banach spaces are provided by Pinelis [256, 257]. A statement for general Banach spaces can be found in [309, Thm. 4]. However, it depends on unspecified constants. Further tail bounds and moment inequalities can be found in the book by Yurinsky [356].

3.2 Risk-Neutral Minimization

We consider the risk-neutral optimal control problem

$$\min_{u\in U_{\mathrm{ad}}} \{\, f(u) = \mathbb{E}[\widehat{J}(u,\xi)] + \Psi(u) \,\}, \tag{3.2.1}$$

where U_{ad} is a convex, closed and nonempty subset of the separable Hilbert space U, $\Psi : U_{\mathrm{ad}} \to \mathbb{R} \cup \{\infty\}$ is convex and lower-semicontinuous, and $\widehat{J} : U \times \Xi \to \mathbb{R}$ is the parameterized cost function. Moreover, $(\Omega, \mathcal{F}, P)$ is a probability space, $(\Xi, \mathcal{F}_\Xi)$ is a measurable space, and $\xi : \Omega \to \Xi$ is measurable. We also use $\xi \in \Xi$ for representing a deterministic element (see p. ix). Let $\xi^1, \xi^2, \ldots$ be independent realizations of ξ such that each ξ^i has the same probability distribution as that of $\xi : \Omega \to \Xi$. We view the random vectors $\xi^i : \Omega^\star \to \Xi$ as mappings defined on a common probability space, which we denote by $(\Omega^\star, \mathcal{F}^\star, P^\star)$; see, e.g., [44, pp. 148–149] for the standard construction of such a space.

The SAA corresponding to (3.2.1) is

$$\min_{u\in U_{\mathrm{ad}}} \{\, f_N(u,\omega) = \mathrm{E}^N[\widehat{J}(u,\xi(\omega))] + \Psi(u) \,\}, \tag{3.2.2}$$

where $\mathrm{E}^N[\widehat{J}(u,\xi(\omega))] = (1/N)\sum_{i=1}^N \widehat{J}(u,\xi^i(\omega))$ for $\omega \in \Omega^\star$. We define $F : U \to \mathbb{R}\cup\{\infty\}$ and the *sample average function* $F_N : U \times \Omega^\star \to \mathbb{R}$ by

$$F(u) = \mathbb{E}[\widehat{J}(u,\xi)] \quad \text{and} \quad F_N(u,\omega) = \mathrm{E}^N[\widehat{J}(u,\xi(\omega))]. \tag{3.2.3}$$

The second argument of f_N and of F_N is often dropped. Throughout the section, we assume that u^* is an optimal solution of (3.2.1) and that $u_N^*(\omega)$ is a minimizer of (3.2.2) for each $\omega \in \Omega^\star$. We refer the reader to [281, Prop. 6.2], [190, Thm. 1], and [192, Prop. 3.12] for theorems on the existence of optimal solutions to stochastic programs.

Let V be a Banach space and let (Ω, F) be a measurable space. A function $\mathsf{f} : V \times \Omega \to \mathbb{R}\cup\{\infty\}$ is a *random lower-semicontinuous function* (or a normal integrand) if, for each $\omega \in \Omega$, $\mathsf{f}(\cdot,\omega)$ is lower-semicontinuous, and f is $\mathcal{B}(V)\otimes\mathsf{F}$-measurable [66, p. 195], [267, pp. 221–222], [294, p. 420]. For example, Carathéodory functions are random lower-semicontinuous [268, p. 175], [66, Lem. III.14]. We recall that $\mathsf{f} : V_1 \times \Omega \to V_2$ is a *Carathéodory mapping* if $\mathsf{f}(\cdot,\omega)$ is continuous for every $\omega \in \Omega$ and $\mathsf{f}(x,\cdot)$ is F-$\mathcal{B}(V_2)$-measurable for all $x \in V_1$ [11, p. 311]. Here, V_1 and V_2 are separable Banach spaces. A Carathéodory function $\mathsf{f} : V_1 \times \Omega \to \mathbb{R}$ is a *convex Carathéodory function* if $\mathsf{f}(\cdot,\omega)$ is convex for all $\omega \in \Omega$.

Assumption 3.2.1. (a) *The set U_{ad} is nonempty, closed, and convex subset of the separable Hilbert space U.*
(b) *The function $F : U \to \mathbb{R}\cup\{\infty\}$ defined in* (3.2.3) *is Gâteaux differentiable at u^*.*
(c) *The penalty function $\Psi : U \to \mathbb{R}\cup\{\infty\}$ is convex and lower-semicontinuous with $\Psi(u) < \infty$ for some $u \in U_{ad}$.*
(d) *The function $\widehat{J} : U \times \Xi \to \mathbb{R}$ is a convex Carathéodory function.*

Conditions on $\widehat{J}$ that ensure the Fréchet differentiability of the function F defined in (3.2.3) are provided, for example, in [126, sect. 4.7], [131, p. A2752], and [128, p. 2079]. Let Assumptions 3.2.1 (a) and 3.2.1 (d) hold. If $\partial F(u^*)$ is a singleton and F is continuous at u^*, then F is Hadamard differentiable at u^* [46, Prop. 2.126] and, hence, Gâteaux differentiable at u^* [46, pp. 34–35].

Assumption 3.2.1 ensures the measurability of the SAA problem's optimal value (see (3.2.2)).

Lemma 3.2.2. *If Assumptions 3.2.1* (a)*, 3.2.1* (c)*, and 3.2.1* (d) *hold, then* $\inf_{u\in U_{ad}} f_N(u,\cdot) : \Omega^\star \to \bar{\mathbb{R}}$ *is measurable, where f_N is defined in* (3.2.2)*. If, in addition,* $\arg\min_{u\in U_{ad}} f_N(u,\omega)$ *is nonempty for all $\omega \in \Omega^\star$, then* $\arg\min_{u\in U_{ad}} f_N(u,\cdot) : \Omega^\star \to U$ *has a measurable selection.*

The proof of Lemma 3.2.2 is provided in section 3.6.1.

3.2.1 Sample Size Estimates for the Optimal Control

Under suitable assumptions, we establish exponential bounds on the tail probabilities of $\|u^* - u_N^*\|_U$, where u^* and $u_N^*(\omega)$ are optimal solutions of (3.2.1) and of (3.2.2), respectively.

Assumption 3.2.3. (a) *The function $\widehat{J}(\cdot,\xi)$ is Gâteaux differentiable on a convex neighborhood of U_{ad} for all $\xi \in \Xi$, and $\nabla_u \widehat{J}(u^*,\cdot) : \Xi \to U$ is measurable.*
(b) *There exists $\alpha > 0$ such that $\widehat{J}(\cdot,\xi)$ is α-strongly convex for each $\xi \in \Xi$.*

We use the following characterization of α-strong convexity: if H is a Hilbert space, $\mathsf{f} : \mathsf{H} \to \mathbb{R} \cup \{\infty\}$ is proper, and $\alpha \geq 0$, then f is α-strongly convex if and only if $\mathsf{f} - (\alpha/2)\|\cdot\|_{\mathsf{H}}^2$ is convex [18, p. 178].
In section 3.2.5, we demonstrate that Assumption 3.2.3 (b) is fulfilled for certain linear-quadratic optimal control problems.

Lemma 3.2.4. *Let Assumptions 3.2.1* (a)*, 3.2.1* (c)*, 3.2.1* (d)*, and 3.2.3* (b) *hold. For each $\omega \in \Omega^\star$, let $u_N^*(\omega)$ be an optimal solution of* (3.2.2)*. Then, for each $\omega \in \Omega^\star$, $u_N^*(\omega)$ is the unique optimal solution of* (3.2.2)*, and $u_N^* : \Omega^\star \to U$ is measurable.*

Proof. The conditions ensure that the objective function $f_N(\cdot, \omega)$ of the SAA problem (3.2.2) is strongly convex for each $\omega \in \Omega^\star$ and that the SAA problem's feasible set U_{ad} is convex. Hence, $u_N^*(\omega)$ is the unique minimizer of (3.2.2) for each fixed $\omega \in \Omega^\star$; see, e.g., [246, p. 48], [46, Lem. 2.33]. Combined with Lemma 3.2.2, we obtain the measurability of u_N^*. □

We impose conditions on the integrability of $\nabla_u \widehat{J}(u^*, \xi) - \nabla F(u^*)$.

Assumption 3.2.5. (a) *For some $\sigma > 0$, we have $\mathbb{E}[\|\nabla_u \widehat{J}(u^*, \xi) - \nabla F(u^*)\|_U^2] \leq \sigma^2$.*
(b) *For some $\tau > 0$, it holds that $\mathbb{E}[\exp(\tau^{-2}\|\nabla_u \widehat{J}(u^*, \xi) - \nabla F(u^*)\|_U^2)] \leq \mathrm{e}$.*

Assumption 3.2.5 (b), when combined with Jensen's inequality, implies Assumption 3.2.5 (a) with $\sigma^2 = \tau^2$; see also [243, p. 1584]. Assumption 3.2.5 (b) and its variants are standard conditions in the literature on stochastic programming [207, eq. (4.1.15)], [99, p. 679], [243, eq. (2.50)], [138, pp. 1035–1036], [294, eq. (5.347)]. For example, if $\|\nabla_u \widehat{J}(u^*, \xi) - \nabla F(u^*)\|_U \leq \rho$ for all $\xi \in \Xi$ and some $\rho > 0$, then Assumption 3.2.5 (b) is satisfied with $\tau = \rho$. In other words, if the U-norm of $\nabla_u \widehat{J}(u^*, \xi) - \nabla F(u^*)$ is essentially bounded or that of $\nabla_u \widehat{J}(u^*, \xi)$, then Assumption 3.2.5 (b) is fulfilled. In section 3.3, we discuss in detail a control problem that satisfies Assumption 3.2.5 (b). Further details on this condition are provided in section 4.2.1.

Exponential Tail Bound

We state our main result, a bound on the tail probabilities of $\|u^* - u_N^*\|_U$.

Theorem 3.2.6. *Let u^* be an optimal solution of* (3.2.1) *and for each $\omega \in \Omega^\star$, let $u_N^*(\omega)$ be a minimizer of* (3.2.2)*. If Assumptions 3.2.1, 3.2.3 and 3.2.5* (b) *hold, then for all $\varepsilon > 0$,*

$$\mathrm{Prob}(\|u^* - u_N^*\|_U \geq \varepsilon) \leq 2\exp(-\tau^{-2} N \varepsilon^2 \alpha^2 / 3). \tag{3.2.4}$$

If, in addition, $\|\nabla_u \widehat{J}(u^*, \xi) - \nabla F(u^*)\|_U \leq \tau$ w.p. 1 (with probability one), then the right-hand side in (3.2.4) *improves to $2\exp(-\tau^{-2} N \varepsilon^2 \alpha^2 / 2)$.*

The proof of Theorem 3.2.6 is presented in section 3.2.2. Theorem 3.2.6 yields a finite sample size estimate.

Remark 3.2.7. Let $\delta \in (0, 1)$ and $\varepsilon > 0$ be arbitrary. We suppose that the hypotheses of Theorem 3.2.6 hold. To obtain $\mathrm{Prob}(\|u^* - u_N^*\|_U \geq \varepsilon) \leq \delta$, we bound the right-hand side in (3.2.4) by δ. Choosing $N \in \mathbb{N}$ with $N \geq \ln(2/\delta)(3\tau^2)/(\varepsilon^2\alpha^2)$ yields $\mathrm{Prob}(\|u^* - u_N^*\|_U \geq \varepsilon) \leq \delta$.

Proposition 3.2.8. *Let u^* be a minimizer of* (3.2.1) *and for each $\omega \in \Omega^\star$, let $u_N^*(\omega)$ be a minimizer of* (3.2.2)*. If Assumptions 3.2.1, 3.2.3 and 3.2.5* (a) *hold, then $\mathbb{E}[\|u^* - u_N^*\|_U^2] \leq \sigma^2/(\alpha^2 N)$.*

The proof of Proposition 3.2.8 is also presented in section 3.2.2. Proposition 3.2.8 and Tschebyshev's inequality imply tail bounds on $\|u^* - u_N^*\|_U$.

Corollary 3.2.9. *If the hypotheses of Proposition 3.2.8 hold, then, for all $\varepsilon > 0$,*

$$\mathrm{Prob}(\|u^* - u_N^*\|_U \geq \varepsilon) \leq \sigma^2/(\varepsilon^2\alpha^2 N). \tag{3.2.5}$$

Corollary 3.2.9 is proven in section 3.2.2.
We compare the tail bound (3.2.4) and (3.2.5). Let $\delta \in (0,1)$ and let $\varepsilon > 0$ be arbitrary, and let the conditions of Corollary 3.2.9 hold. The sample size estimate $N \geq \sigma^2/(\varepsilon^2\alpha^2\delta)$ implies $\mathrm{Prob}(\|u^* - u_N^*\|_U \geq \varepsilon) \leq \delta$. This estimate depends linearly on $1/\delta$ in contrast to that provided by Remark 3.2.7.
We demonstrate that the dependence of the tail bound (3.2.4) on the problem's data τ, α is essentially optimal for the problem class modeled by Assumptions 3.2.1, 3.2.3 and 3.2.5 (b). We verify the optimality by constructing an explicit model problem that satisfies Assumptions 3.2.1, 3.2.3 and 3.2.5 (b). For this model problem, the tail bound (3.2.5) turns out to be conservative as a function of the accuracy $\varepsilon > 0$.

Example 3.2.10. The following example is inspired by that in Shapiro [292, Ex. 1]; see also [294, Ex. 5.21]. We consider

$$\min_{u \in H_0^1(\mathcal{D})} (\alpha/2)\|u\|^2_{H_0^1(\mathcal{D})} - \mathbb{E}[(b(\xi), u)_{L^2(\mathcal{D})}], \tag{3.2.6}$$

where $\alpha > 0$, $\mathcal{D} = (0,1)$, and $b : \mathbb{R}^2 \to L^2(\mathcal{D})$ is defined by $b(\xi)(x) = \pi^2\xi_1\varphi_1(x) + 4\pi^2\xi_2\varphi_2(x)$. Here, ξ_1, $\xi_2 : \Omega \to \mathbb{R}$ are independent, mean-zero Gaussian random variables with unit variance, and φ_1, $\varphi_2 : \mathcal{D} \to \mathbb{R}$ are given by

$$\varphi_1(x) = (\sqrt{2}/\pi)\sin(\pi x) \quad \text{and} \quad \varphi_2(x) = (\sqrt{2}/(2\pi))\sin(2\pi x). \tag{3.2.7}$$

The space $H_0^1(\mathcal{D})$ is equipped with the norm $\|\cdot\|_{H_0^1(\mathcal{D})} = \|\mathrm{D}\cdot\|_{L^2(\mathcal{D})}$ (see p. viii).
Since $\mathbb{E}[b(\xi)] = 0$, the optimal solution u^* to (3.2.6) is $u^* = 0$. The SAA problem of (3.2.6) is

$$\min_{u \in H_0^1(\mathcal{D})} (\alpha/2)\|u\|^2_{H_0^1(\mathcal{D})} - (\mathrm{E}^N[b(\xi(\omega))], u)_{L^2(\mathcal{D})}. \tag{3.2.8}$$

Its optimal solution $u_N^*(\omega)$ for $\omega \in \Omega^\star$ is characterized by the canonical optimality condition of (3.2.8). This condition is the elliptic PDE $(u_N^*(\omega)', v')_{L^2(\mathcal{D})} = (1/\alpha)(\mathrm{E}^N[b(\xi(\omega))], v)_{L^2(\mathcal{D})}$ for all $v \in H_0^1(\mathcal{D})$ [54, pp. 58 and 67]. We have $u_N^* = (1/\alpha)\mathrm{E}^N[\xi_1]\varphi_1 + (1/\alpha)\mathrm{E}^N[\xi_2]\varphi_2$.
Below, we show that Assumption 3.2.5 (b) is satisfied with $\tau^2 = 4/(1 - \exp(-2))$, and that

$$\mathrm{Prob}(\|u_N^* - u^*\|_{H_0^1(\mathcal{D})} \geq \varepsilon) = \exp(-N\alpha^2\varepsilon^2/2) \quad \text{for all} \quad \varepsilon > 0. \tag{3.2.9}$$

This tail bound reveals that the exponential order of the tail bound in (3.2.4) is optimal up to the constant $3\tau^2/2 \approx 6.9$.
Let us establish (3.2.9). Using (3.2.7), we obtain $(\varphi_1, \varphi_2)_{H_0^1(\mathcal{D})} = 0$, $\|\varphi_1\|_{H_0^1(\mathcal{D})} = 1$, and $\|\varphi_2\|_{H_0^1(\mathcal{D})} = 1$. We conclude that $\|a_1\varphi_1 + a_2\varphi_2\|^2_{H_0^1(\mathcal{D})} = a_1^2 + a_2^2$ for all a_1, $a_2 \in \mathbb{R}$. In particular, $\|u_N^*\|^2_{H_0^1(\mathcal{D})} = (1/\alpha)^2\mathrm{E}^N[\xi_1]^2 + (1/\alpha)^2\mathrm{E}^N[\xi_2]^2$. Since $\mathrm{E}^N[\xi_1] \sim \mathcal{N}(0, N^{-1}\alpha^{-2})$ and $\mathrm{E}^N[\xi_2] \sim \mathcal{N}(0, N^{-1}\alpha^{-2})$ are independent, the distribution of $N\alpha^2\|u_N^*\|^2_{H_0^1(\mathcal{D})}$ is the chi-squared distribution χ_2^2 with two degrees of freedom [86, p. 13]. Hence

$$\mathrm{Prob}(\|u_N^*\|_{H_0^1(\mathcal{D})} \geq \varepsilon) = \mathrm{Prob}(N\alpha^2\|u_N^*\|^2_{H_0^1(\mathcal{D})} \geq N\alpha^2\varepsilon^2) = \mathrm{Prob}(\chi_2^2 \geq N\alpha^2\varepsilon^2) = \mathrm{e}^{-N\alpha^2\varepsilon^2/2}$$

for all $\varepsilon \geq 0$ [86, p. 13]. Combined with $\|u_N^* - u^*\|_{H_0^1(\mathcal{D})} = \|u_N^*\|_{H_0^1(\mathcal{D})}$, we obtain (3.2.9).

We verify Assumptions 3.2.1 and 3.2.3. We fix $\xi \in \mathbb{R}^2$. For this example, we have

$$\Psi = 0, \quad \widehat{J}(u,\xi) = (\alpha/2)\|u\|^2_{H^1_0(\mathcal{D})} + (b(\xi),u)_{L^2(\mathcal{D})} \quad \text{and} \quad F(u) = (\alpha/2)\|u\|^2_{H^1_0(\mathcal{D})}.$$

Let us define $B : \mathbb{R}^2 \to H^1_0(\mathcal{D})^*$ by $\langle B(\xi), v\rangle_{H^1_0(\mathcal{D})^*, H^1_0(\mathcal{D})} = (b(\xi),u)_{L^2(\mathcal{D})}$. The operator $B(\xi)$ is linear and bounded [151, p. 30], $\widehat{J}(\cdot,\xi)$ is α-strongly convex, and

$$\|\mathrm{D}_u\widehat{J}(u^*,\xi) - \mathrm{D}F(u^*)\|_{H^1_0(\mathcal{D})^*} = \|B(\xi)\|_{H^1_0(\mathcal{D})^*}. \tag{3.2.10}$$

Putting together the pieces, we conclude that Assumptions 3.2.1 and 3.2.3 are satisfied.
To verify Assumption 3.2.5 (b), we compute $\|B(\xi)\|_{H^1_0(\mathcal{D})^*}$. Let us define $y : \mathbb{R}^2 \to H^1_0(\mathcal{D})$ by $y(\xi) = \xi_1\varphi_1 + \xi_2\varphi_2$, and fix $\xi \in \mathbb{R}^2$. Since φ_1 and φ_2 are orthonormal in $H^1_0(\mathcal{D})$, we have $\|y(\xi)'\|^2_{L^2(\mathcal{D})} = \|y(\xi)\|^2_{H^1_0(\mathcal{D})} = \xi_1^2 + \xi_2^2$. It holds that $-y(\xi)'' = b(\xi)$. Hence, $y(\xi)$ is the Riesz representation of $B(\xi)$ in $(H^1_0(\mathcal{D}), \|\cdot\|_{H^1_0(\mathcal{D})})$ (see, e.g., [151, p. 28 and Thm. 1.4]) which implies

$$\|B(\xi)\|_{H^1_0(\mathcal{D})^*} = \|y(\xi)\|_{H^1_0(\mathcal{D})} = (\xi_1^2 + \xi_2^2)^{1/2}. \tag{3.2.11}$$

We show that $\mathbb{E}[\exp(\tau^{-2}\|B(\xi)\|^2_{H^1_0(\mathcal{D})^*})] \le \mathrm{e}$ for $\tau^2 = 4/(1-\exp(-2))$. In light of (3.2.10), this estimate implies Assumption 3.2.5 (b), when identifying $H^1_0(\mathcal{D})^*$ with $H^1_0(\mathcal{D})$. We have $\mathbb{E}[\exp(s\xi_i^2/2)] = \mathrm{e}$ for $i = 1, 2$, and $s = 1 - \exp(-2) \in (0,1)$ [57, p. 9]. Combined with the independence of ξ_1 and ξ_2, Jensen's inequality, and (3.2.11), we obtain

$$\mathbb{E}[\exp(\tau^{-2}\|B(\xi)\|^2_{H^1_0(\mathcal{D})^*})] = \mathbb{E}[\mathrm{e}^{\xi_1^2/\tau^2}]\mathbb{E}[\mathrm{e}^{\xi_2^2/\tau^2}] = \mathbb{E}[\mathrm{e}^{s\xi_1^2/4}]\mathbb{E}[\mathrm{e}^{s\xi_2^2/4}] \le \mathrm{e}^{1/2}\mathrm{e}^{1/2} = \mathrm{e}.$$

For later reference, we compute additional characteristics of (3.2.6) and of (3.2.8). We have $f_N(u_N^*) = (\alpha/2)\|u_N^*\|^2_U - \alpha\|u_N^*\|^2_U$. Moreover, it holds that $u^* = 0$, $\|u_N^*\|^2_{H^1_0(\mathcal{D})} = (1/\alpha)^2\mathrm{E}^N[\xi_1]^2 + (1/\alpha)^2\mathrm{E}^N[\xi_2]^2$ and $\mathbb{E}[b(\xi)] = 0$. Combined with $\xi_1, \xi_2 \in \mathcal{N}(0,1)$, we conclude that

$$\mathbb{E}[\|u_N^* - u^*\|^2_{H^1_0(\mathcal{D})}] = \tfrac{2}{\alpha N}, \quad \mathbb{E}[f(u^*)] = 0, \quad \mathbb{E}[f(u_N^*)] = \tfrac{1}{\alpha N}, \quad \mathbb{E}[f_N(u_N^*)] = -\tfrac{1}{\alpha N}.$$

If $\alpha = 0$, then the set of optimal solutions of (3.2.6) is $H^1_0(\mathcal{D})$ because $\mathbb{E}[b(\xi)] = 0$. In this case, the corresponding SAA problem (3.2.8) has no optimal solution w.p. 1 because its objective function is linear and $\mathrm{E}^N[b(\xi)] \neq 0$ w.p. 1. Indeed, using (3.2.11), the $H^1_0(\mathcal{D})$-orthogonality of φ_1, φ_2, and ξ_1, $\xi_2 \sim \mathcal{N}(0,1)$, we find that $\|\mathrm{E}^N[B(\xi)]\|^2_{H^1_0(\mathcal{D})^*} = \mathrm{E}^N[\xi_1]^2 + \mathrm{E}^N[\xi_2]^2 \neq 0$ w.p. 1.
It may be possible to derive similar tail bounds than that in (3.2.9) when b is a non-truncated, Hilbert space-valued, Gaussian random variable. However, the derivation of exact bounds or of tight lower bounds is more involved; see Yurinsky [356, Thm. 2.3.1, and sects. 2.3.3 and 2.3.4].

3.2.2 Proof of Sample Size Estimates for the Optimal Control

We prove Theorem 3.2.6, Proposition 3.2.8, and Corollary 3.2.9 using Lemmas 3.2.11–3.2.15 and Theorem 3.2.16.
Bochner integrable subgradients of a convex random lower-semicontinuous function are unbiased estimators of the expectation function's Gâteaux derivative; see, e.g., [138, p. 1050]. We use this fact to deduce that $\nabla_u\widehat{J}(u^*,\xi)$ is an unbiased estimator for the gradient of the function $F = \mathbb{E}[\widehat{J}(\cdot,\xi)]$ defined in (3.2.3) at the optimal solution u^* of (3.2.1).

Lemma 3.2.11. *Let $(\Omega, \mathsf{F}, \mathsf{P})$ be a probability space, and let V be Banach space. Let $\mathsf{f} : V \times \Omega \to \mathbb{R} \cup \{\infty\}$ be random lower-semicontinuous and $\mathsf{f}(\cdot,\omega)$ be convex for all $\omega \in \Omega$. Suppose that $\mathsf{F} : V \to \mathbb{R} \cup \{\infty\}$ defined by $\mathsf{F}(x) = \int_\Omega \mathsf{f}(x,\omega)\mathrm{d}\mathsf{P}(\omega)$ is well-defined and Gâteaux differentiable at $x \in \operatorname{dom}\mathsf{F}$, the domain of F. Let $\mathsf{g} : \Omega \to V^*$ be Bochner integrable with $\mathsf{g}(\omega) \in \partial_x\mathsf{f}(x,\omega)$ for almost every $\omega \in \Omega$. Then $\mathrm{D}\mathsf{F}(x) = \int_\Omega \mathsf{g}(\omega)\mathrm{d}\mathsf{P}(\omega)$.*

Proof. Fix $h \in V$. We show that $\mathsf{F}(x+th) \in \mathbb{R}$ for all sufficiently small $t > 0$. We have $\mathsf{F}(x) \in \mathbb{R}$ and $\mathrm{D}\mathsf{F}(x)[h] \in \mathbb{R}$. Let us define $\phi : (0,\infty) \to \mathbb{R} \cup \{\infty\}$ by $\phi(t) = t^{-1}(\mathsf{F}(x+th) - \mathsf{F}(x))$. Using $\mathbb{R} \ni \mathsf{F}'(x;h) = \inf_{t>0} \phi(t)$ [46, p. 49], we obtain that, for all $\varepsilon > 0$, there exists $t > 0$ with $\phi(t) \leq \mathsf{F}'(x;h) + \varepsilon$. Because ϕ is increasing [46, p. 49], we have $\mathsf{F}(x+th) \in \mathbb{R}$ for all sufficiently small $t > 0$.
We establish $\mathrm{D}\mathsf{F}(x)[h] = \int_\Omega \mathsf{g}(\omega)[h]\mathrm{d}\mathsf{P}(\omega)$. Fix $t > 0$ with $\mathsf{F}(x+th) \in \mathbb{R}$. For almost every $\omega \in \Omega$, we have $\mathsf{f}(x+th,\omega) - \mathsf{f}(x,\omega) \geq t\mathsf{g}(\omega)[h]$. Since g is Bochner integrable and $\mathsf{F}(x+th) \in \mathbb{R}$, we obtain $\mathsf{F}(x+th) - \mathsf{F}(x) \geq t\int_\Omega \mathsf{g}(\omega)[h]\mathrm{d}\mathsf{P}(\omega)$. Therefore, $\mathrm{D}\mathsf{F}(x)[h] \geq \int_\Omega \mathsf{g}(\omega)[h]\mathrm{d}\mathsf{P}(\omega)$ and, hence, $\mathrm{D}\mathsf{F}(x)[w] = \int_\Omega \mathsf{g}(\omega)[w]\mathrm{d}\mathsf{P}(\omega)$ for all $w \in V$. Combined with the Bochner integrability of g, we obtain $\mathrm{D}\mathsf{F}(x) = \int_\Omega \mathsf{g}(\omega)\mathrm{d}\mathsf{P}(\omega)$ [36, p. 78]. □

Lemma 3.2.12. *Let $(\Omega, \mathsf{F}, \mathsf{P})$ be a probability space, H be a real, separable Hilbert space and let $\mathsf{R} : \mathsf{H} \to \mathsf{H}^*$ be the Riesz mapping defined by $\langle \mathsf{R}[x], z\rangle_{\mathsf{H}^*,\mathsf{H}} = (x,z)_\mathsf{H}$ for all $z \in \mathsf{H}$. Then, the measurability (integrability) of $\mathsf{g} : \Omega \to \mathsf{H}^*$ implies that of $\mathsf{R}^{-1}\mathsf{g} : \Omega \to \mathsf{H}$, and the measurability (integrability) of $\mathsf{h} : \Omega \to \mathsf{H}$ implies that of $\mathsf{R}\mathsf{h} : \Omega \to \mathsf{H}^*$.*

Proof. The Riesz representation theorem [151, Thm. 1.4] ensures that R is bijective, isometric and linear. Hence, the measurability (integrability) of g and h implies that of $\mathsf{R}^{-1}\mathsf{g}$ and $\mathsf{R}\mathsf{h}$, respectively; see, e.g., [169, Lem. 1.5]. □

Lemma 3.2.13. *Let Assumptions 3.2.1 and 3.2.3 (a) hold. Suppose that $\widehat{J}(\cdot,\xi)$ is α-strongly convex for all $\xi \in \Xi$ and some $\alpha \geq 0$. Then, the function F_N defined in (3.2.3) is Gâteaux differentiable on a convex neighborhood of U_{ad}, $\mathbb{E}[\nabla_u \widehat{J}(u^*,\xi)] = \nabla F(u^*)$, and w.p. 1,*

$$(\nabla F_N(u_2) - \nabla F_N(u_1), u_2 - u_1)_U \geq \alpha\|u_2 - u_1\|_U^2 \quad \text{for all} \quad u_1, u_2 \in U_{ad}. \tag{3.2.12}$$

Proof. Assumption 3.2.1 (d), the sum rule, and the definition of F_N (see (3.2.3)) imply its Gâteaux differentiability on a convex neighborhood V of U_{ad}. Since, for each $\xi \in \Xi$, $\widehat{J}(\cdot,\xi)$ is α-strongly convex, and F_N is Gâteaux differentiable on V, we obtain (3.2.12) [246, p. 48]. Assumption 3.2.3 (a), when combined with Jensen's inequality, implies that $\nabla_u \widehat{J}(u^*,\xi)$ is (Bochner) integrable since

$$(\mathbb{E}[\|\nabla_u \widehat{J}(u^*,\xi)\|_U])^2 \leq 2\mathbb{E}[\|\nabla_u \widehat{J}(u^*,\xi) - \nabla F(u^*)\|_U^2] + 2\|\nabla F(u^*)\|_U^2 < \infty. \tag{3.2.13}$$

Combined with Lemmas 3.2.11 and 3.2.12 and the fact that U is a real, separable Hilbert space (see p. viii and Assumption 3.2.1 (a)), we conclude that $\mathbb{E}[\nabla_u \widehat{J}(u^*,\xi)] = \nabla F(u^*)$. □

We establish necessary optimality conditions for (3.2.1) and its SAA problem (3.2.2).

Lemma 3.2.14. *Let Assumptions 3.2.1 and 3.2.3 (a) hold, and let $\omega \in \Omega^\star$ be arbitrary. Suppose that $u^* \in U_{ad}$ is an optimal solution of (3.2.1) and that $u_N^* = u_N^*(\omega) \in U_{ad}$ is a minimizer of (3.2.2). Then, for all $u \in U_{ad}$,*

$$\begin{aligned} (\nabla F(u^*), u - u^*)_U + \Psi(u) - \Psi(u^*) &\geq 0, \\ (\nabla F_N(u_N^*), u - u_N^*)_U + \Psi(u) - \Psi(u_N^*) &\geq 0, \end{aligned} \tag{3.2.14}$$

and $(\nabla F_N(u_N^*) - \nabla F(u^*), u^* - u_N^*)_U \geq 0$, where F and F_N are defined in (3.2.3).

Proof. Lemma 3.2.13 implies that the sample average function $F_N : U \to \mathbb{R}$ is Gâteaux differentiable at u_N^* and convex (see Assumptions 3.2.3 (a) and 3.2.1 (d)). Moreover, F is Gâteaux differentiable at u^* and convex by Assumptions 3.2.1 (c) and 3.2.1 (d). The set U_{ad} is convex (see Assumption 3.2.1 (a)), and Ψ is proper, convex and lower-semicontinuous (see Assumption 3.2.1 (c)). Now, the proof of (3.2.14) follows from that of [163, Thm. 3.1] by Ito

and Kunisch [163]. We have $\Psi(u^*)$, $\Psi(u_N^*) \in \mathbb{R}$. Choosing $u = u_N^*$ in the first inequality in (3.2.14) and $u = u^*$ in the second estimate, and adding the resulting inequalities yields $(\nabla F_N(u_N^*) - \nabla F(u^*), u^* - u_N^*)_U \geq 0$. □

We combine Lemma 3.2.13 with the optimality conditions stated in Lemma 3.2.14 to obtain an error estimate for the SAA problem's optimal control.

Lemma 3.2.15. *If the hypotheses of Theorem 3.2.6 hold, then w.p.* 1*,*

$$\alpha\|u^* - u_N^*\|_U \leq \|\nabla F_N(u^*) - \nabla F(u^*)\|_U, \tag{3.2.15}$$

where F and F_N are defined in (3.2.3)*.*

Proof. Choosing $u_2 = u^*$ and $u_1 = u_N^*$ in (3.2.12), we find that

$$(\nabla F_N(u^*) - \nabla F_N(u_N^*), u^* - u_N^*)_U \geq \alpha\|u^* - u_N^*\|_U^2. \tag{3.2.16}$$

Lemma 3.2.14 gives $(\nabla F_N(u_N^*) - \nabla F(u^*), u^* - u_N^*)_U \geq 0$. Combined with (3.2.16) and the Cauchy–Schwarz inequality, we conclude that

$$\begin{aligned}\alpha\|u^* - u_N^*\|_U^2 &\leq (\nabla F_N(u^*) - \nabla F_N(u_N^*), u^* - u_N^*)_U + (\nabla F_N(u_N^*) - \nabla F(u^*), u^* - u_N^*)_U \\ &= (\nabla F_N(u^*) - \nabla F(u^*), u^* - u_N^*)_U \\ &\leq \|\nabla F_N(u^*) - \nabla F(u^*)\|_U \|u^* - u_N^*\|_U.\end{aligned}$$ □

The estimate (3.2.15) implies that an accurate estimation of $\nabla F_N(u^*)$ yields an accurate approximation u_N^* of the optimal solution u^* to (3.2.1), provided that $\alpha > 0$.
We state large deviations results for sums of independent Hilbert space-valued random variables. The large deviation results and the estimate (3.2.15) are used to establish the exponential tail bound provided in Theorem 3.2.6.

Theorem 3.2.16. *Let $(\Omega, \mathsf{F}, \mathsf{P})$ be a probability space and let H be a separable Hilbert space. Suppose that $\mathsf{Z}_i : \Omega \to \mathsf{H}$ for $i = 1, 2, \ldots$ are independent, mean-zero random variables such that $\mathbb{E}[\exp(\tau^{-2}\|\mathsf{Z}_i\|_{\mathsf{H}}^2)] \leq \mathrm{e}$ for some $\tau > 0$. Then, for each $N \in \mathbb{N}$ and every $\varepsilon \geq 0$,*

$$\mathrm{Prob}(\|\mathsf{S}_N/N\|_{\mathsf{H}} \geq \varepsilon) \leq 2\exp(-\tau^{-2}\varepsilon^2 N/3),$$

where $\mathsf{S}_N = \mathsf{Z}_1 + \cdots + \mathsf{Z}_N$. If, in addition, $\|\mathsf{Z}_i\|_{\mathsf{H}} \leq \tau$ w.p. 1 for $i = 1, 2, \ldots$, then $\mathrm{Prob}(\|\mathsf{S}_N/N\|_{\mathsf{H}} \geq \varepsilon) \leq 2\exp(-\tau^{-2}\varepsilon^2 N/2)$.

Proof. The proof is presented in section 3.6.2. □

Proof of Theorem 3.2.6. Lemma 3.2.4 ensures the measurability of $u_N^* : \Omega^\star \to U$, where for each $\omega \in \Omega^\star$, $u_N^*(\omega)$ is an optimal solution of the SAA problem (3.2.2). For each $\varepsilon > 0$, Lemma 3.2.15 gives

$$\mathrm{Prob}(\|u^* - u_N^*\|_U \geq \varepsilon) \leq \mathrm{Prob}(\|\nabla F_N(u^*) - \nabla F(u^*)\|_U \geq \varepsilon\alpha).$$

Using (3.2.3), Assumption 3.2.1 (b), and Lemma 3.2.13, we get

$$\nabla F_N(u^*) - \nabla F(u^*) = (1/N)\sum_{i=1}^{N}(\nabla_u \widehat{J}(u^*, \xi^i) - \nabla F(u^*)). \tag{3.2.17}$$

Lemma 3.2.13 and the fact that ξ^i have the same distribution as ξ imply $\mathbb{E}[\nabla_u \widehat{J}(u^*, \xi^i)] = \nabla F(u^*)$ for $i = 1, \ldots, N$. Moreover, the independence of ξ^i and the measurability of $\nabla_u \widehat{J}(u^*, \cdot)$ (see Assumption 3.2.3 (a)) imply that $\nabla_u \widehat{J}(u^*, \xi^i)$ are independent [44, p. 399]. Hence, the U-valued random variables $\nabla_u \widehat{J}(u^*, \xi^i) - \nabla F(u^*)$ have zero mean and are independent. Combined with the separability of U (see Assumption 3.2.1 (a)) and Theorem 3.2.16, we obtain (3.2.4). □

Proof of Proposition 3.2.8. Lemma 3.2.4 implies that $u_N^* : \Omega^\star \to U$ is measurable. Lemma 3.2.13 yields $\mathbb{E}[\nabla_u \widehat{J}(u^*, \xi)] = \nabla F(u^*)$. Combined with Lemma 3.2.15, the fact that U is a Hilbert space (see Assumption 3.2.1 (a)), the independence of ξ^i and (3.2.17), we conclude that

$$\mathbb{E}[\|u^* - u_N^*\|_U^2] \leq \tfrac{1}{\alpha^2}\mathbb{E}[\|\nabla F_N(u^*) - \nabla F(u^*)\|_U^2] = \tfrac{1}{N\alpha^2}\mathbb{E}[\|\nabla_u \widehat{J}(u^*, \xi) - \nabla F(u^*)\|_U^2].$$

□

Proof of Corollary 3.2.9. Proposition 3.2.8 gives $\mathbb{E}[\|u^* - u_N^*\|_U^2] \leq \sigma^2/(\alpha^2 N)$. Hence, Tschebyshev's inequality yields $\text{Prob}(\|u^* - u_N^*\|_U \geq \varepsilon) \leq \sigma^2/(\alpha^2 \varepsilon^2 N)$ for each $\varepsilon > 0$. □

3.2.3 Discussion

We compare the bounds provided by Theorem 3.2.6 and Proposition 3.2.8 with similar estimates from the literature.

Our approach for deriving the tail bound (3.2.4) can be interpreted as an adaption of that by Shapiro [287, 288] for nonlinear, finite-dimensional stochastic optimization problems to the problem class modeled by Assumptions 3.2.1, 3.2.3 and 3.2.5 (b).

Kouri and Shapiro [190] establish

$$\alpha\|u^* - u_N^*\|_U \leq \|\nabla F_N(u_N^*) - \nabla F(u_N^*)\|_U, \tag{3.2.18}$$

[190, eq. (42)], assuming $\widehat{J}(\cdot, \xi) : U \to \mathbb{R}$ is continuously differentiable for each $\xi \in \Xi$, $\Psi = 0$, and the function F defined in (3.2.3) is α-strongly convex with $\alpha > 0$. Here, $u_N^*(\omega)$ is an optimal solution of (3.2.2) for $\omega \in \Omega^\star$, and u^* is a minimizer of (3.2.1). In contrast to the estimate (3.2.15), the right-hand side in (3.2.18) depends on the random control u_N^*. This dependence stops us from applying the tail bound provided by Theorem 3.2.16. However, the convexity assumption on F is weaker than that imposed by Assumption 3.2.3 (b), which requires the integrand $\widehat{J}(\cdot, \xi)$ to be α-strongly convex for all $\xi \in \Xi$.

For stochastic approximation, Geiersbach and Pflug [128] establish a bound on $\mathbb{E}[\|u_k - u^*\|_U^2]$ that decreases like $1/k$ as the iteration counter k increases, and depends on $1/\alpha^2$ in a similar way as the bound provided by Proposition 3.2.8. For the derivation of their bound, Geiersbach and Pflug [128] require that the expectation function F defined in (3.2.1) is α-strongly convex and that $\mathbb{E}[\|\mathsf{G}(u, \xi)\|_U^2] \leq M_1 < \infty$ for all $u \in U_{\text{ad}}$ [128, pp. 2083 and 2087]. Here, u_k $(k = 1, 2, \ldots)$ are the iterates of the projected stochastic gradient method [128, Alg. 2.1], and $\mathsf{G} : U \times \Xi \to U$ is the stochastic gradient, such as $\mathsf{G} = \nabla_u \widehat{J}$. Similar bounds on $\mathbb{E}[\|u_k - u^*\|_U^2]$ are provided in [243, eq. (2.9)] and [131, Thm. 3.2], where u_k are the iterates of stochastic approximation methods.

The estimate (3.2.15) is similar to that established by Vexler [326, Prop. 3.5] for the variational discretization of a linear-quadratic control problem. Whereas the estimate in [326, Prop. 3.5] is deterministic, both the finite element approximation and the SAA approach yield perturbed optimization problems. It is therefore not surprising that similar techniques can be used for some parts of the perturbation analysis. However, the perturbation analysis of the sampling error (3.2.15) differs from the error analysis of the variational discretization.

3.2.4 Confidence Bounds for the Optimal Value

We provide non-asymptotic bounds on the optimal value of the SAA problem (3.2.1), that is, bounds on $f_N(u_N^*) - f(u^*)$ as a function of N, where $f(u^*)$ is the optimal value of (3.1.1), and $f_N(u_N^*)$ is that of its SAA (3.1.2). Moreover, u^* is an optimal solution of (3.1.1) and, for each $\omega \in \Omega^\star$, $u_N^*(\omega)$ is a minimizer of (3.1.2). Our derivation is built on that performed by Guigues,

Juditsky, and Nemirovski [138] for optimization problems posed in $(\mathbb{R}^n, \|\cdot\|)$ "equipped" with a distance-generating function.
We consider convex optimization problems with essentially bounded objective functions and gradient mappings.

Assumption 3.2.17. (a) *For some $\tau_1 > 0$, we have $|\widehat{J}(u^*, \xi) - F(u^*)| \le \tau_1$ for all $\xi \in \Xi$.*
(b) *For some $\tau_2 > 0$, it holds that $\|\nabla_u \widehat{J}(u^*, \xi) - \nabla F(u^*)\|_U \le \tau_2$ for all $\xi \in \Xi$.*

We derive tail bounds on $f_N(u_N^*) - f(u^*)$, the difference of the optimal values of (3.2.2) and of its SAA problem (3.2.1).

Proposition 3.2.18. *Let u^* be a minimizer of* (3.2.1) *and for each $\omega \in \Omega^\star$, let $u_N^*(\omega)$ be a minimizer of* (3.2.2)*. Let Assumptions 3.2.1, 3.2.3* (a) *and 3.2.17 hold. Then, for all $\varepsilon > 0$,*

$$\operatorname{Prob}(f_N(u_N^*) > f(u^*) + \varepsilon) \le \exp(-\varepsilon^2 \tau_1^{-2} N/2). \tag{3.2.19}$$

If, in addition, $R_{ad} = \sup_{u \in U_{ad}} \|u - u^*\|_U < \infty$, then, for all ε_1, $\varepsilon_2 > 0$,

$$\operatorname{Prob}(f_N(u_N^*) + \varepsilon_1 + \varepsilon_2 < f(u^*)) \le \exp(-\varepsilon_1^2 \tau_1^{-2} N/2) + 2\exp(-\varepsilon_2^2 \tau_2^{-2} R_{ad}^{-2} N/2), \tag{3.2.20}$$

where f is defined in (3.2.1) *and f_N in* (3.2.2)*.*

The bound (3.2.19) is a consequence of Hoeffding's bound. If $Z_i : \Omega^\star \to [a_i, b_i]$ are independent and a_i, $b_i \in \mathbb{R}$ $(i = 1, \ldots, N)$, then Hoeffding's bound [152, Thm. 2] gives, for all $\varepsilon > 0$,

$$\operatorname{Prob}\left(\frac{1}{N}\sum_{i=1}^{N}(Z_i - \mathbb{E}[Z_i]) \ge \varepsilon\right) \le \mathrm{e}^{-\frac{2N^2\varepsilon^2}{\sum_{i=1}^{N}(b_i - a_i)^2}}.$$

Proof of Proposition 3.2.18. The proof is inspired by that of [138, Prop. 1]. Lemma 3.2.2 implies that $f_N(u_N^*)$ is measurable. Consequently, the events in (3.2.19) and (3.2.20) are well-defined. We prove (3.2.19). Using the definition of f_N (see (3.2.2)), and that of F_N and of F (see (3.2.3)), we obtain

$$f_N(u_N^*) \le f_N(u^*) = F_N(u^*) + \Psi(u^*) = F_N(u^*) - F(u^*) + f(u^*). \tag{3.2.21}$$

We define the mean-zero random variables $Z_i = \widehat{J}(u^*, \xi^i) - \mathbb{E}[\widehat{J}(u^*, \xi^i)]$. Using (3.2.3), we obtain $F_N(u^*) - F(u^*) = (1/N)\sum_{i=1}^{N} Z_i$. Owing to Assumption 3.2.17 and the independence of ξ^i, we can apply Hoeffding's bound to $F_N(u^*) - F(u^*) = (1/N)\sum_{i=1}^{N}(Z_i - \mathbb{E}[Z_i])$ which yields (3.2.19). We establish (3.2.20). Since F_N is Gâteaux differentiable at u^* (see Lemma 3.2.13) and convex (see Assumption 3.2.1 (d)), we have (see, e.g., [46, Prop. 2.125])

$$F_N(u_N^*) - F_N(u^*) \ge (\nabla F_N(u^*), u_N^* - u^*)_U. \tag{3.2.22}$$

The optimality condition (3.2.14) yields $(\nabla F(u^*), u_N^* - u^*)_U + \Psi(u_N^*) - \Psi(u^*) \ge 0$. Combined with (3.2.22), the Cauchy–Schwarz inequality, and $\|u_N^* - u^*\|_U \le R_{\mathrm{ad}}$, we find that

$$\begin{aligned} f_N(u_N^*) = F_N(u_N^*) + \Psi(u_N^*) &\ge F_N(u^*) + \Psi(u_N^*) + (\nabla F_N(u^*), u_N^* - u^*)_U \\ &\ge F_N(u^*) + \Psi(u^*) + (\nabla F_N(u^*) - \nabla F(u^*), u_N^* - u^*)_U \\ &\ge f(u^*) + F_N(u^*) - F(u^*) - \|\nabla F_N(u^*) - \nabla F(u^*)\|_U R_{\mathrm{ad}}. \end{aligned} \tag{3.2.23}$$

Assumption 3.2.17 allows us to apply Hoeffding's bound to $F(u^*) - F_N(u^*)$ which yields

$$\operatorname{Prob}(F(u^*) - F_N(u^*) > \varepsilon_1) \le \exp(-N\tau_1^{-2}\varepsilon_1^2/2). \tag{3.2.24}$$

Lemma 3.2.13 gives $\mathbb{E}[\nabla_u \widehat{J}(u^*,\xi)] = \nabla F(u^*)$. Thus, Assumption 3.2.17 and (3.2.17) also allow us to apply Theorem 3.2.16 to $R_{\mathrm{ad}}\|\nabla F_N(u^*) - \nabla F(u^*)\|_U$. We obtain

$$\mathrm{Prob}(R_{\mathrm{ad}}\|\nabla F_N(u^*) - \nabla F(u^*)\|_U > \varepsilon_2) \leq 2\exp(-\varepsilon_2^2\tau_2^{-2}R_{\mathrm{ad}}^{-2}N/2).$$

Combined with (3.2.24) and the union bound/Boole's inequality, we obtain (3.2.20). □

Proposition 3.2.19. *Let Assumptions 3.2.1, 3.2.3* (a) *and 3.2.5* (a) *hold. If, in addition,* $R_{ad} = \sup_{u\in U_{ad}} \|u - u^*\|_U < \infty$*, then*

$$\mathbb{E}[f_N(u_N^*)] \leq f(u^*) \leq \mathbb{E}[f_N(u_N^*)] + (R_{ad}/\sqrt{N})\sigma, \tag{3.2.25}$$

where F *is defined in* (3.2.3)*,* f *in* (3.2.1)*, and* f_N *in* (3.2.2)*.*

Proof. Taking expectations of (3.2.21), we get $\mathbb{E}[f_N(u_N^*)] \leq f(u^*)$. Since $\mathbb{E}[\|\nabla_u \widehat{J}(u^*,\xi) - \nabla F(u^*)\|_U^2] \leq \sigma^2$ and U is a Hilbert space, we obtain $\mathbb{E}[\|\nabla F_N(u^*) - \nabla F(u^*)\|_U^2] \leq \sigma^2/N$. Jensen's inequality ensures $\mathbb{E}[\|\nabla F_N(u^*) - \nabla F(u^*)\|_U] \leq \sigma/\sqrt{N}$. Combined with (3.2.23), we obtain $f(u^*) \leq \mathbb{E}[f_N(u_N^*)] + (R_{\mathrm{ad}}/\sqrt{N})\sigma$. □

Proposition 3.2.20. *If Assumptions 3.2.1, 3.2.3* (a) *and 3.2.5* (a) *hold, then*

$$\mathbb{E}[f_N(u_N^*)] \leq f(u^*) \leq \mathbb{E}[f_N(u_N^*)] + \sigma^2/(2\alpha N).$$

Proof. The lower bound follows from (3.2.21). To establish the upper bound, we use the fact that $\widehat{J}(\cdot,\xi)$ is α-strongly convex for all $\xi \in \Xi$. As opposed to (3.2.23), we obtain

$$\begin{aligned} f_N(u_N^*) &\geq F_N(u^*) + \Psi(u_N^*) + (\nabla F_N(u^*), u_N^* - u^*)_U + (\alpha/2)\|u_N^* - u^*\|_U^2 \\ &\geq F_N(u^*) + \Psi(u^*) + (\nabla F_N(u^*) - \nabla F(u^*), u_N^* - u^*)_U + (\alpha/2)\|u_N^* - u^*\|_U^2 \\ &\geq f(u^*) + F_N(u^*) - F(u^*) - (1/(2\alpha))\|\nabla F_N(u^*) - \nabla F(u^*)\|_U^2, \end{aligned}$$

where we have used the Cauchy–Schwarz inequality and Young's inequality to get $2|(\nabla F_N(u^*) - \nabla F(u^*), u_N^* - u^*)_U| \leq (1/\alpha)\|\nabla F_N(u^*) - \nabla F(u^*)\|_U^2 + \alpha\|u_N^* - u^*\|_U^2$. Taking expectations, we obtain the upper bound. □

Discussion

Proposition 3.2.18 yields confidence bounds on $f(u^*)$; cf. [138, pp. 1036–1037]. Let $\varepsilon > 0$ and $\delta \in (0,1)$ be arbitrary, and let the hypotheses of Proposition 3.2.18 be fulfilled. If $N \geq 2\ln(4/\delta)(\tau_1 + R_{\mathrm{ad}}\tau_2)^2/\varepsilon^2$, then Proposition 3.2.18 ensures with $c = \tau_1/(\tau_2 R_{\mathrm{ad}})$,

$$\mathrm{Prob}(f(u^*) \in [f_N(u_N^*) - \varepsilon c/(c+1), f_N(u_N^*) + \varepsilon]) \geq 1 - \delta. \tag{3.2.26}$$

To verify this bound, we define $\varepsilon_1 = c\varepsilon_2$ and $\varepsilon_2 = \varepsilon/(c+1)$. We have $\varepsilon = \varepsilon_1 + \varepsilon_2$, $\varepsilon_1 = \varepsilon c/(c+1)$ and $\varepsilon_1^2\tau_1^{-2} = c^2\varepsilon_2^2\tau_1^{-2} = \varepsilon_2^2\tau_2^{-2}R_{\mathrm{ad}}^{-2}$. Combined with Proposition 3.2.18 and the union bound, we find that

$$\mathrm{Prob}(f(u^*) + \varepsilon_1 < f_N(u_N^*) < f(u^*) - \varepsilon_1 - \varepsilon_2) \leq 2\mathrm{e}^{-\varepsilon_1^2\tau_1^{-2}N/2} + 2\mathrm{e}^{-\varepsilon_2^2\tau_2^{-2}R_{\mathrm{ad}}^{-2}N/2} = 4\mathrm{e}^{-\varepsilon_1^2\tau_1^{-2}N/2}.$$

It must yet be shown that the above condition on N is the same as $N \geq 2\ln(4/\delta)\tau_1^2/\varepsilon_1^2$, which ensures $4\exp(-\varepsilon_1^2\tau_1^{-2}N/2) \leq \delta$. Using $\varepsilon_1 = c\varepsilon_2$, $\varepsilon_2 = \varepsilon/(c+1)$ and $c = \tau_1/(\tau_2 R_{\mathrm{ad}})$, we find that

$$\tau_1^2/\varepsilon_1^2 = \tau_1^2/(c^2\varepsilon_2^2) = \tau_1^2(c+1)^2/(c^2\varepsilon^2) = \tau_2^2 R_{\mathrm{ad}}^2(c+1)^2/\varepsilon^2 = (\tau_1 + R_{\mathrm{ad}}\tau_2)^2/\varepsilon^2.$$

Therefore, the above condition on N is the same as $N \geq 2\ln(4/\delta)\tau_1^2/\varepsilon_1^2$. Putting together the pieces, we obtain (3.2.26).

In addition to the lower bound on the optimal value $f(u^*)$ provided by Proposition 3.2.19, we have $\mathbb{E}[\inf_{u\in U_{\mathrm{ad}}} f_N(u)] \leq \mathbb{E}[\inf_{u\in U_{\mathrm{ad}}} f_{N+1}(u)] \leq \inf_{u\in U_{\mathrm{ad}}} f(u)$ [294, Prop. 5.6], valid without convexity assumptions about $\widehat{J}(\cdot,\xi)$ for $\xi \in \Xi$. In our setting, we have $f(u^*) = \inf_{u\in U_{\mathrm{ad}}} f(u)$ and $\mathbb{E}[\inf_{u\in U_{\mathrm{ad}}} f_N(u)] = \mathbb{E}[f_N(u_N^*)]$ because we assume the existence of optimal solutions for both (3.2.1) and its SAA (3.2.2). These inequalities assert that the SAA problem's optimal value is a downward biased estimator of the "true" optimal objective function value, and the bias decreases monotonically as the sample size increases; see also [294, p. 171]. The tail bound (3.2.19) asserts that $f_N(u_N^*)$ is to some extent concentrated below $f(u^*)$. The bound (3.2.19) is also valid without convexity assumptions because it solely relies on the estimate (3.2.21), provided that $\widehat{J}(u^*,\cdot)$ is essentially bounded or more generally sub-Gaussian.

Example 3.2.10 shows that the bound provided by Proposition 3.2.20 is optimal.

As opposed to Proposition 3.2.19, stochastic approximation yields the estimate $\mathbb{E}[f(\bar{u}_1^K)] \leq f(u^*) + (D_{\mathrm{ad}}M/\sqrt{K})$, when used with iterate averaging, constant step size, and a fixed number of iterations K. Here, $D_{\mathrm{ad}} = \sup_{u\in U_{\mathrm{ad}}} \|u - u_1\|_U$, $\mathbb{E}[\|\mathsf{G}(u,\xi)\|_U^2] \leq M^2 < \infty$ for all $u \in U_{\mathrm{ad}}$, and $\mathsf{G} : U \times \Xi \to U$ is the stochastic gradient, such as $\mathsf{G} = \nabla_u \widehat{J}$. Moreover, $u_1 \in U_{\mathrm{ad}}$ is the (deterministic) initial value, and $\bar{u}_1^K \in U_{\mathrm{ad}}$ is a weighted average of the iterates. See [243, eqns. (2.5), (2.17), and (2.21)], [131, Thm. 3.3], [128, pp. 2088], and [246, p. 192].

For stochastic approximation, Nemirovski, Juditsky, Lan, and Shapiro [243, Prop. 2.2] establish exponential tail bounds on $f(\bar{u}_1^K) > f(u^*)$, when $(U, \|\cdot\|_U) = (\mathbb{R}^n, \|\cdot\|)$ is "equipped" with a distance-generating function. We conjecture that similar bounds can be established when U is an infinite-dimensional (separable) Hilbert space. In this case, tail bounds for $\|\bar{u}_1^K - u^*\|_U$ can be established under suitable assumptions: if f is α-strongly convex with $\alpha > 0$ and subdifferentiable at u^*, $\bar{u}_1^K \in U_{\mathrm{ad}}$, and u^* is an optimal solution of (3.2.1), then $(\alpha/2)\|\bar{u}_1^K - u^*\|_U^2 \leq f(\bar{u}_1^K) - f(u^*)$. We conclude that $\mathrm{Prob}(\|\bar{u}_1^K - u^*\|_U > (2\varepsilon/\alpha)^{1/2}) \leq \mathrm{Prob}(f(\bar{u}_1^K) > f(u^*) + \varepsilon)$ for $\varepsilon \geq 0$.

If the hypotheses of Proposition 3.2.8 hold, $\Psi = 0$, u^* is an unconstrained minimizer of (3.2.1), and ∇F is Lipschitz continuous with Lipschitz constant $L > 0$, then we have

$$\mathbb{E}[f(u_N^*)] \leq f(u^*) + (L/2)\mathbb{E}[\|u_N^* - u^*\|_U^2] \leq f(u^*) + \sigma^2 L/(2\alpha^2 N),$$

cf. [243, eqns. (2.12) and (2.13)], [131, eq. (3.11)]. These conditions are fulfilled for Example 3.2.10 with $L = \alpha$, and Example 3.2.10 shows that these estimates are essentially optimal.

3.2.5 Application to Linear-Quadratic Optimal Control under Uncertainty

We consider the linear-quadratic optimal control problem with convex regularization

$$\min_{u\in U_{\mathrm{ad}}} \{\, (1/2)\mathbb{E}[\|Q(\xi)S(u,\xi) - y_d\|_H^2] + (\alpha/2)\|u\|_U^2 + \Psi(u) \,\}, \tag{3.2.27}$$

where $\alpha \geq 0$, $S : U \times \Xi \to Y$ is the parameterized solution operator of the affine-linear operator equation (3.2.29), and $Q : \Xi \to \mathscr{L}(Y,H)$. Moreover, $y_d \in H$ and H is a Hilbert space. In this section, U and U_{ad}, and $\Psi : U \to \mathbb{R}\cup\{\infty\}$ fulfill Assumptions 3.2.1 (a) and 3.2.1 (c), respectively. We refer the reader to [151, sect. 1.5.1] for the formulation of a general deterministic linear-quadratic control problem with control constraints. We can model parameterized affine-linear elliptic and parabolic PDEs with (3.2.29), such as the heat equation with random inputs considered in [230, sect. 3.1.2]. Moreover, the optimization problems with affine-linear elliptic PDEs considered in [122, sect. 7], [228, Chap. 3], [227, sect. 2], [191, sect. 6.1], [184, sect. 3.4], [190, sect. 6], [139], and [308] can be formulated as instances of (3.2.27).

If $\mathcal{D} \subset \mathbb{R}^d$ is a bounded domain and $U = L^2(\mathcal{D})$, we can choose $\Psi = \gamma\|\cdot\|_{L^1(\mathcal{D})}$ for $\gamma \geq 0$. This nonsmooth regularization has been considered, for example, in [303], [335], [91, sect. 6.1], and [321, sect. 9.3] for deterministic optimal control problems, and in [129, sect. 4] for a risk-neutral control problem with a semilinear elliptic PDE. Further nonsmooth regularizers are considered in [163, sect. 5].
We define $J : Y \times U \times \Xi \to \mathbb{R}$ and the reduced parameterized cost function $\widehat{J} : U \times \Xi \to \mathbb{R}$ by

$$J(y, u, \xi) = (1/2)\|Q(\xi)y - y_d\|_H^2 + (\alpha/2)\|u\|_U^2 \quad \text{and} \quad \widehat{J}(u, \xi) = J(S(u, \xi), u, \xi). \tag{3.2.28}$$

Following [159, Def. 1.1.27] and [36, Def. 2.23], we refer to an operator $\mathsf{G} : \Xi \to \mathscr{L}(V_1, V_2)$ as *strongly measurable* (w.r.t. the strong operator topology) if $\mathsf{G}(\cdot)x : \Xi \to V_2$ is strongly measurable for each $x \in V_2$. Here, V_1 and V_2 are Banach spaces.

Assumption 3.2.21. (a) *The spaces Y and Z are separable Banach spaces, and U and H are separable Hilbert spaces.*
(b) *The operators $A : \Xi \to \mathscr{L}(Y, Z)$, $B : \Xi \to \mathscr{L}(U, Z)$ and $Q : \Xi \to \mathscr{L}(Y, H)$, and $g : \Xi \to Z$ are strongly measurable, and $y_d \in H$.*
(c) *For each $\xi \in \Xi$, $A(\xi)$ has a bounded inverse.*

We consider deterministic desired states $y_d \in H$, but it is also possible to consider random ones [121, p. 27], [124, p. 4]. Instead of assuming the strong measurability of the mappings A, B, Q and g, it would be sufficient to impose their strong $\mathbb{P}$-measurability, where $\mathbb{P}$ is the probability distribution of $\xi : \Omega \to \Xi$.
Let Assumption 3.2.21 hold. Let us define the parameterized solution operator $S : U \times \Xi \to Y$: for each $(u, \xi) \in U \times \Xi$, $S(u, \xi)$ is the solution to

$$\text{Find} \quad y \in Y : \quad A(\xi)y + B(\xi)u = g(\xi). \tag{3.2.29}$$

Owing to Assumption 3.2.21, we have for each $(u, \xi) \in U \times \Xi$,

$$S(u, \xi) = A(\xi)^{-1}[g(\xi) - B(\xi)u]. \tag{3.2.30}$$

As in [151, sect. 1.6.3], we identify U^* and H^* with U and H, respectively. We define $z : U \times \Xi \to Z^*$ as the solution corresponding to the parameterized adjoint equation

$$\text{Find} \quad z \in Y : \quad A(\xi)^* z = -Q(\xi)^*[Q(\xi)S(u, \xi) - y_d], \tag{3.2.31}$$

where S is defined in (3.2.30) and $\xi \in \Xi$. Assumption 3.2.21 ensures that the parameterized adjoint equation (3.2.31) has a unique solution for each $\xi \in \Xi$ [196, pp. 49 and 236]. Combined with the computations in [151, sects. 1.6.2 and 1.6.3], we find that the gradient of the function $\widehat{J}(\cdot, \xi)$ defined in (3.2.28) is

$$\nabla_u \widehat{J}(u, \xi) = B(\xi)^* z(u, \xi) + \alpha u. \tag{3.2.32}$$

Lemma 3.2.22. *If Assumption 3.2.21 holds, then the following mappings are Carathéodory:* (a) *$S : U \times \Xi \to Y$ defined in* (3.2.30), (b) *$\widehat{J} : U \times \Xi \to \mathbb{R}$ defined in* (3.2.28), *and* (c) *$\nabla_u \widehat{J} : U \times \Xi \to U$ defined in* (3.2.32).

Proof. (a) We show that S (see (3.2.30)) is a Carathéodory mapping. Fix $\xi \in \Xi$. The mapping $S(\cdot, \xi)$ is affine-linear, and $A(\xi)^{-1}$ and $B(\xi)$ are bounded. Hence, $S(\cdot, \xi)$ is continuous.
Now, fix $u \in U$. The strong measurability of $A : \Xi \to \mathscr{L}(Y, Z)$ implies that of $A^{-1} : \Xi \to \mathscr{L}(Z, Y)$ [36, Thms. 2.15 and 2.16]. Hence, [159, Prop. 1.1.28] implies that $A^{-1}(\cdot)g(\cdot)$ is strongly measurable, and [159, Cor. 1.1.29] yields the strong measurability of $A^{-1}(\cdot)B(\cdot)u$. Consequently,

$S(u,\cdot)$ is strongly measurable, as it is the sum of strongly measurable mappings [150, Thm. 3.5.4]. Putting together the pieces, we conclude that S is a Carathéodory mapping.
(b) We establish that $\widehat{J}$ (see (3.2.28)) is a Carathéodory function. Fix $\xi \in \Xi$. By (3.2.28), we have $\widehat{J}(\cdot,\xi) = J(S(\cdot,\xi),\cdot,\xi)$. The continuity of $J(\cdot,\cdot,\xi)$ and of $S(\cdot,\xi)$ imply that of $\widehat{J}(\cdot,\xi)$.
It must yet be shown that $\widehat{J}(u,\cdot)$ is $\mathcal{F}_\Xi$-measurable for each $u \in U$. Fix $u \in U$. We define $G : \Xi \to Y \times U$ by $G(\xi) = (S(u,\xi),u)$. Using part (a) and [66, Lem. III.14], we obtain that $S(u,\cdot)$ is $\mathcal{F}_\Xi$-$\mathcal{B}(Y)$-measurable. Combined with [169, Lems. 1.5 and 1.8], we find that G is $\mathcal{F}_\Xi$-$\mathcal{B}(Y) \otimes \mathcal{B}(U)$-measurable. Since $\widehat{J}(u,\xi) = J(\cdot,\cdot,\xi) \circ G(\xi)$ for each $\xi \in \Xi$ and the function $J : (Y \times U) \times \Xi \to \mathbb{R}$ defined in (3.2.28) is a Carathéodory function, [11, Lem. 8.2.3] and [169, Lem. 1.2] ensure that $\widehat{J}(u,\cdot)$ is $\mathcal{F}_\Xi$-measurable.
(c) We show that $\nabla_u \widehat{J}(u,\xi)$ (see (3.2.32)) is a Carathéodory mapping. Owing to (3.2.31), we have $z(u,\xi) = -A(\xi)^{-*}Q(\xi)^*(Q(\xi)S(u,\xi) - y_d)$. Using this formula and part (a), we obtain that $B(\xi)^* z(\cdot,\xi)$ is continuous for each $\xi \in \Xi$. According to [36, Thm. 2.16], A^{-*}, B^* and Q^* are strongly measurable. Now, we can establish the measurability of $z(u,\cdot)$ using the same arguments as in part (a). □

Lemma 3.2.23. *Let Assumption 3.2.21 hold. We define $F_1 : U \to \mathbb{R}$ by*

$$F_1(u) = (1/2)\mathbb{E}[\|Q(\xi)S(u,\xi) - y_d\|_H^2],$$

and $K : \Xi \to \mathscr{L}(U,H)$ by $K(\xi) = -Q(\xi)A(\xi)^{-1}B(\xi)$. If, in addition, $\mathbb{E}[\|B(\xi)^* z(u,\xi)\|_U] < \infty$ and $\mathbb{E}[\|K(\xi)^*K(\xi)u\|_U] < \infty$ for all $u \in U$, then F_1 is infinitely times continuously differentiable with $\nabla F_1(u) = \mathbb{E}[B(\xi)^* z(u,\xi)]$ and $\nabla^2 F_1(u)[v] = \mathbb{E}[K(\xi)^*K(\xi)v]$ for all u, $v \in U$. Here, S is defined in (3.2.30) *and z in* (3.2.31).

Proof. Fix u, v, $w \in U$ and $\xi \in \Xi$. We define $\widehat{J}_1 : U \times \Xi \to \mathbb{R}$ by $\widehat{J}_1(u,\xi) = \widehat{J}(u,\xi) - (\alpha/2)\|u\|_U^2$, where $\widehat{J}$ is defined in (3.2.28). Using (3.2.32) and the definition of z provided in (3.2.31), we have $\nabla_u \widehat{J}_1(u,\xi) = B(\xi)^* z(u,\xi)$. Combined with $\mathbb{E}[\|B(\xi)^* z(u,\xi)\|_U] < \infty$, we find that

$$\mathbb{E}[(\nabla_u \widehat{J}_1(u,\xi),v)_U] = \mathbb{E}[(B(\xi)^* z(u,\xi),v)_U] = (\mathbb{E}[B(\xi)^* z(u,\xi)],v)_U$$

and that $v \mapsto (\mathbb{E}[B(\xi)^* z(u,\xi)],v)_U$ is a bounded linear operator. Formally, we obtain $\nabla F_1(u) = \mathbb{E}[B(\xi)^* z(u,\xi)]$. Since $S(\cdot,\xi)$ is affine linear (see (3.2.30)), we have $S(u+v,\xi) - S(u,\xi) = S_u(u,\xi)[v]$ and $S_{uu}(u,\xi) = 0$. Combined with the formulas for second derivatives provided in [151, sect. 1.6.5] and the definition of K, we obtain

$$\begin{aligned}(\nabla_{uu}\widehat{J}_1(u,\xi)w,v)_U &= \langle Q(\xi)^*Q(\xi)S_u(u,\xi)w, S_u(u,\xi)v\rangle_{Y^*,Y} = (Q(\xi)S_u(u,\xi)w, Q(\xi)S_u(u,\xi)v)_H \\ &= (K(\xi)w, K(\xi)v)_H = (w, K(\xi)^*K(\xi)v)_U.\end{aligned}$$

Since $\mathbb{E}[\|K(\xi)^*K(\xi)v\|_U] < \infty$ for all $v \in U$, the operator $v \mapsto \mathbb{E}[K(\xi)^*K(\xi)v]$ is linear and bounded [150, Thm. 3.8.2]. Formally, we obtain $\nabla^2 F_1(u)[v] = \mathbb{E}[K(\xi)^*K(\xi)v]$. Putting together the pieces and using the fact that $\widehat{J}_1(\cdot,\xi)$ is quadratic for all $\xi \in \Xi$, we conclude that F_1 is twice Gâteaux differentiable and, hence, infinitely many times continuously differentiable. □

Examples

Many instances of the linear-quadratic control problem (3.2.27) frequently encountered in the literature are defined by the following data: $\alpha > 0$, $\Psi = 0$, $H = U$, $Q \in \mathscr{L}(Y,H)$ is the (deterministic) embedding operator of the embedding $Y \hookrightarrow U$, and $B \in \mathscr{L}(U,Z)$ and $g \in Z$ are deterministic.[1] Moreover, $\Psi = 0$, U_{ad} is nonempty, closed and convex, and $A : \Xi \to \mathscr{L}(Y,Z)$ is

[1] Parameterized affine-linear elliptic state equations of the type (3.2.29) where $g : \Xi \to Z$ is random can be found, for example, in [125, sect. 2.2], [190, sect. 6], and [192, sect. 4].

strongly measurable and there exist constants $0 < \kappa_{\min}^* \leq \kappa_{\max}^*$ with $\kappa_{\min}^* \|y\|_Y \leq \|A(\xi)y\|_Z \leq \kappa_{\max}^* \|y\|_Y$ for all $(y,\xi) \in Y \times \Xi$. See [131, p. A2758], [227, sect. 2], [122, p. 20], [121, p. 31], for example. Since $\kappa_{\min}^* \|y\|_Y \leq \|A(\xi)y\|_Z \leq \kappa_{\max}^* \|y\|_Y$ for all $(y,\xi) \in Y \times \Xi$, the operator $A(\xi)$ has a bounded inverse for each $\xi \in \Xi$ with $\|A(\xi)^{-1}z\|_Y \leq (1/\kappa_{\min}^*)\|z\|_Z$ for all $(z,\xi) \in Z \times \Xi$ [196, p. 101]. We conclude that Assumption 3.2.21 holds true. Lemmas 3.2.22 and 3.2.23, when combined with that fact that the mappings $\widehat{J}(u,\cdot)$ and $\nabla_u \widehat{J}(u,\cdot)$ (see (3.2.32)) are essentially bounded, imply that Assumptions 3.2.1, 3.2.3, 3.2.5 and 3.2.17 are satisfied, provided that (3.2.27) has an optimal solution.
We verify the strong measurability of a random elliptic operator A using Lemma 3.2.24.

Lemma 3.2.24. *If $\mathcal{D} \subset \mathbb{R}^d$ is a bounded domain, then $\phi : L^\infty(\mathcal{D}) \to \mathscr{L}(H_0^1(\mathcal{D}), H_0^1(\mathcal{D})^*)$ defined by $\langle \phi(\kappa)y, v\rangle_{H_0^1(\mathcal{D})^*, H_0^1(\mathcal{D})} = (\kappa \nabla y, \nabla v)_{L^2(\mathcal{D})^d}$ is Lipschitz continuous with Lipschitz constant one.*

Proof. The mapping ϕ is well-defined [151, pp. 29–30]. Since $\|\cdot\|_{H_0^1(\mathcal{D})} = |\cdot|_{H^1(\mathcal{D})}$ (see p. viii), Hölder's inequality ensures $|\langle \phi(\kappa_2) - \phi(\kappa_1)y, v\rangle_{Y^*,Y}| = |((\kappa_2 - \kappa_1)\nabla y, \nabla v)_{L^2(\mathcal{D})^d}| \leq \|\kappa_2 - \kappa_1\|_{L^\infty(\mathcal{D})}$ for all y, $v \in H_0^1(\mathcal{D})$ with norm one. Hence, ϕ is Lipschitz continuous with constant one. □

We define $A : \Xi \to \mathscr{L}(H_0^1(\mathcal{D}), H_0^1(\mathcal{D})^*)$ by $\langle A(\xi)y, v\rangle_{H_0^1(\mathcal{D})^*, H_0^1(\mathcal{D})} = (\kappa(\xi)\nabla y, \nabla v)_{L^2(\mathcal{D})^d}$ with $\kappa : \Xi \to L^\infty(\mathcal{D})$ being strongly measurable—a common example for A in the literature on PDE-constrained optimization under uncertainty; see, e.g., [131, sect. 4], [226, sect. 2.1] and [230, sects. 3.1.1 and 4.1]. Here, $\mathcal{D} \subset \mathbb{R}^d$ is a bounded domain. For each $\xi \in \Xi$, we indeed have $A(\xi) \in \mathscr{L}(H_0^1(\mathcal{D}), H_0^1(\mathcal{D})^*)$ [151, pp. 29–30]. We show that A is strongly measurable w.r.t. the uniform operator topology which implies that A is strongly measurable (w.r.t. the strong operator topology) [159, p. 12]. Since the function ϕ defined in Lemma 3.2.24 is continuous, it is measurable [169, Lem. 1.5]. Combined with $A = \phi \circ \kappa$ and the fact that κ is strongly measurable, we find that A is strongly measurable w.r.t. the uniform operator topology [159, Cor. 1.1.11].
For $\kappa \in L^\infty(\mathcal{D} \times \Xi)$, the operator $A : \Xi \to \mathscr{L}(H_0^1(\mathcal{D}), H_0^1(\mathcal{D})^*)$ can also be defined by $\langle A(\xi)y, v\rangle_{H_0^1(\mathcal{D})^*, H_0^1(\mathcal{D})} = (\kappa(\cdot,\xi)\nabla y, \nabla v)_{L^2(\mathcal{D})^d}$; see, e.g., [122, sect. 7], [124, sect. 2.1], [227, sect. 2]. Here, $\mathcal{D} \subset \mathbb{R}^d$ is a bounded domain and $L^\infty(\mathcal{D} \times \Xi)$ is the Bochner space of essentially bounded, real-valued functions w.r.t. the product of the Lebesgue measure on $\mathcal{D}$ and of the probability distribution $\mathbb{P}$ of ξ. The operator A is strongly $\mathbb{P}$-measurable w.r.t. the uniform operator topology [124, sect. 2.1].
Choosing κ as a log-normal random diffusion coefficients, that is, $\ln(\kappa)$ is a Banach space-valued Gaussian random variable, has been popular in the literature [12, 3, 69]. In this case, Assumption 3.2.5 (b) is generally violated. We construct an explicit example using Lemma 3.2.25, which is inspired by [303, Lem. 3.1] established by Stadler [303].

Lemma 3.2.25. *Consider the linear-quadratic control problem* (3.2.27) *with $U = L^2(\mathcal{D})$, $\alpha \geq 0$, and $\Psi = \gamma\|\cdot\|_{L^1(\mathcal{D})}$, where $\mathcal{D} \subset \mathbb{R}^d$ is a bounded domain and $\gamma \geq 0$. Suppose that the hypotheses of Lemma 3.2.23 hold and that U_{ad} is convex with $0 \in U_{ad}$. If $\|\mathbb{E}[B(\xi)^* z(0,\xi)]\|_{L^\infty(\mathcal{D})} \leq \gamma < \infty$, then the zero function is a minimizer of* (3.2.27)*, where B is defined in Assumption 3.2.21, and z in* (3.2.31)*. If $\alpha > 0$ in* (3.2.27)*, then the zero function is the minimizer of* (3.2.27)*.*

Proof. We show that $0 \in \partial[F(0) + \Psi(0)]$, where $F + \Psi$ is the convex cost function of (3.2.27) (see (3.2.3)). Lemma 3.2.23, the identification of U^* with U, and the (Lipschitz) continuity of Ψ [1, Thm. 2.8] ensure $\partial[F(0) + \Psi(0)] = \nabla F(0) + \partial\Psi(0)$ [46, Prop. 2.125 and Thm. 2.168]. Combined with $\nabla F(0) = \mathbb{E}[B(\xi)^* z(0,\xi)]$ (see Lemma 3.2.23) and $\partial\Psi(0) = \{\, \lambda \in L^2(\mathcal{D}) : |\lambda| \leq \gamma \,\}$ [303, eq. (2.3)], we obtain $0 \in \partial[F(0) + \Psi(0)]$. (The inequality $|\lambda| \leq \gamma$ is understood in a pointwise almost everywhere sense for $\lambda \in L^2(\mathcal{D})$.) If $\alpha > 0$, then the objective function of (3.2.27) is α-strongly convex. Hence, its minimum over the convex set U_{ad} is unique [246, p. 48]. □

We construct an elliptic control problem with a log-normal random diffusion coefficient for which Assumption 3.2.5 (b) is violated. We choose

$$\begin{aligned} &Y = H_0^1(\mathcal{D}), \quad Z = Y^*, && U = L^2(\mathcal{D}) = U_{\mathrm{ad}} = H, && \Xi = \mathbb{R}, \quad \mathcal{D} = (0,1), \\ &Q(\xi)y = y, && y_d = 2, && \Psi = \gamma\|\cdot\|_{L^1(\mathcal{D})}, && \gamma \geq 0, \quad \alpha > 0, \\ &B(\xi)u = -u, && g = 0, && \langle A(\xi)y, v\rangle_{Y^*,Y} = (\exp(-\xi)y', v')_{L^2(\mathcal{D})}. \end{aligned}$$

Moreover, $\xi : \Omega \to \mathbb{R}$ is a standard Gaussian random variable. Since $\mathbb{E}[\exp(\xi^2/4)] < \infty$ [57, p. 9], we deduce that the hypotheses of Lemma 3.2.23 are satisfied. Using (3.2.30), we get $S(0,\xi) = 0$ for all $\xi \in \mathbb{R}$. Combined with (3.2.31), we find that $z(0,\xi)(x) = \exp(\xi)x(1-x)$ for all $x \in [0,1]$. Thus, $\|\mathbb{E}[z(0,\xi)]\|_{L^\infty(\mathcal{D})}$ is finite. We choose $\gamma = \|\mathbb{E}[z(0,\xi)]\|_{L^\infty(\mathcal{D})}$. For this data, the minimizer of (3.2.27) is the zero function according to Lemma 3.2.25. We have $\|\cdot(1-\cdot)\|^2_{L^2(\mathcal{D})} = 1/30$. Using (3.2.32) and Lemma 3.2.23 yields

$$\|\nabla_u \widehat{J}(0,\xi) - \nabla F(0)\|_U = \|z(0,\xi) - \mathbb{E}[z(0,\xi)]\|_U = (1/30)(\exp(\xi) - \mathbb{E}[\exp(\xi)]),$$

where $\widehat{J}$ is defined in (3.2.28) and F in (3.2.3). Combined with the fact that $\mathbb{E}[\exp(\xi^2/2)] = \infty$ [57, p. 9], we obtain that, for all $\tau \in \mathbb{R}_{++}$, $\mathbb{E}[\exp(\tau^{-2}\|\nabla_u \widehat{J}(0,\xi) - \nabla F(0)\|_U^2)] = \infty$. Thus, Assumption 3.2.5 (b) is violated.

Discussion

The objective function of (3.2.27) has some "hidden" composite structure. Let the hypotheses of Lemma 3.2.23 be fulfilled. Using [159, eq. (1.2)], we obtain, for all $u \in U$,

$$\mathbb{E}[(\mathbb{E}[Q(\xi)S(u,\xi)] - y_d, Q(\xi)S(u,\xi) - \mathbb{E}[Q(\xi)S(u,\xi)])_H] = 0.$$

Consequently, we have, for all $u \in U$,

$$\mathbb{E}[\|Q(\xi)S(u,\xi) - y_d\|_H^2] = \|\mathbb{E}[Q(\xi)S(u,\xi)] - y_d\|_H^2 + \mathbb{E}[\|Q(\xi)S(u,\xi) - \mathbb{E}[Q(\xi)S(u,\xi)]\|_H^2];$$

see also [325, pp. 176–177]. The first term is the squared of the average distance of $Q(\xi)S(u,\xi)$ to y_d, and the latter is the strong centered second moment of $Q(\xi)S(u,\xi)$.
The control problem (3.2.27) can be simplified under a mild condition on the data. We define the modified solution operator $\widetilde{S} : U \times \Xi \to Y$ and the modified adjoint state $\widetilde{z} : U \times \Xi \to Z^*$ by

$$\widetilde{S}(u,\xi) = A^{-1}(\xi)[\mathbb{E}[g(\xi)] - B(\xi)u] \quad \text{and} \quad \widetilde{z}(u,\xi) = -A^{-*}(\xi)Q^*(\xi)[Q(\xi)\widetilde{S}(u,\xi) - y_d],$$

cf. (3.2.30) and (3.2.31). Let Assumption 3.2.21 hold. If g and the modified adjoint state $\widetilde{z}(u,\cdot)$ are independent (w.r.t. the probability distribution of ξ) and Bochner integrable for all $u \in U$, then the linear-quadratic problem (3.2.27) is equivalent to

$$\min_{u \in U_{\mathrm{ad}}} \{\, (1/2)\mathbb{E}[\|Q(\xi)\widetilde{S}(u,\xi) - y_d\|_H^2] + (\alpha/2)\|u\|_U^2 + \Psi(u) \,\}. \tag{3.2.33}$$

To verify the equivalence, let us fix $(u,\xi) \in U \times \Xi$. Using (3.2.30), we have $\widetilde{S}(u,\xi) = S(u,\xi) - w(\xi)$ for $w(\xi) = A^{-1}(\xi)g(\xi) - A^{-1}(\xi)\mathbb{E}[g(\xi)]$. We recall that U^* and H^* are identified with U and H, respectively. Consequently, we have

$$\begin{aligned} (Q(\xi)w(\xi), Q(\xi)\widetilde{S}(u,\xi) - y_d)_H &= \langle Q^*(\xi)[Q(\xi)\widetilde{S}(u,\xi) - y_d], w(\xi)\rangle_{Y^*,Y} \\ &= \langle \widetilde{z}(u,\xi), g(\xi) - \mathbb{E}[g(\xi)]\rangle_{Z^*,Z}. \end{aligned}$$

Combined with independence assumption on g and $\widetilde{z}(u,\cdot)$, and [160, Prop. 6.1.3], we obtain the identity $\mathbb{E}[(Q(\xi)w(\xi), Q(\xi)\widetilde{S}(u,\xi) - y_d)_H] = 0$. Putting together the pieces, we conclude that

$$\begin{aligned}\mathbb{E}[\|Q(\xi)S(u,\xi) - y_d\|_H^2] &= \mathbb{E}[\|Q(\xi)[\widetilde{S}(u,\xi) - w(\xi)] - y_d\|_H^2] \\ &= \mathbb{E}[\|Q(\xi)\widetilde{S}(u,\xi) - y_d\|_H^2] + \mathbb{E}[\|Q(\xi)w(\xi)\|_H^2].\end{aligned}$$

Because $\mathbb{E}[\|Q(\xi)w(\xi)\|_H^2]$ is independent of the control $u \in U$, we obtain that (3.2.27) and (3.2.33) are equivalent.

We show that the function $F_1 : U \to \mathbb{R}$ defined by $F_1(u) = (1/2)\mathbb{E}[\|Q(\xi)S(u,\xi) - y_d\|_H^2]$ is not strongly convex under the following conditions:

- The hypotheses of Lemma 3.2.23 hold, the embedding $Y \hookrightarrow H$ is compact, and $\dim(H) = \infty$.
- Let $\iota \in \mathscr{L}(H, Y)$ be the embedding operator of the compact embedding $Y \hookrightarrow H$. We choose $H = U$, and define $Q(\xi) = \iota$ and $K : \Xi \to \mathscr{L}(U,U)$ by $K(\xi) = -\iota A(\xi)^{-1}B(\xi)$.
- We suppose that K is strongly measurable w.r.t. the uniform operator topology and that $\mathbb{E}[\|K(\xi)\|_{\mathscr{L}(U,U)}^2] < \infty$.

Below, we show that these hypotheses imply that $\nabla^2 F_1$ is not coercive which is equivalent to the fact that F_1 is not strongly convex. This may suggest that the strong convexity of the objective function of (3.2.27) solely comes from the control regularizer $(\alpha/2)\|\cdot\|_U^2$.

We have $\mathbb{E}[\|K(\xi)\|_{\mathscr{L}(U,U)}^2] = \mathbb{E}[\|K(\xi)^*K(\xi)\|_{\mathscr{L}(U,U)}]$ [196, Thm. 3.9-4]. Here, $K(\xi)^*$ is the Hilbert space-adjoint of $K(\xi)$ for $\xi \in \Xi$. We define $T : \Xi \to \mathscr{L}(U,U)$ by $T(\xi) = K(\xi)^*K(\xi)$. The strong measurability of K w.r.t. the uniform operator topology implies that of K^* [36, Thm. 2.16]. We deduce that T is strongly measurable w.r.t. the uniform operator topology [159, Cor. 1.1.29]. By assumption, $K(\xi) = -\iota A(\xi)^{-1}B(\xi)$ is compact [196, p. 411] and, therefore, $T(\xi)$ is compact [196, p. 411]. Now, we show that $\mathbb{E}[T(\xi)]$ is the uniform limit of compact operators. We consider the sample mean $\mathrm{E}^N[T(\xi)](\omega) = (1/N)\sum_{i=1}^N T(\xi^i(\omega))$, where $\xi^i : \Omega^\star \to \Xi$ are independent with the same distribution as $\xi : \Omega \to \Xi$. (The probability space $(\Omega^\star, \mathcal{F}^\star, P^\star)$ is as in section 3.2.) Since $\mathscr{L}(U,U)$ is a Banach space [196, Thm. 2.10-2], $T(\xi^i) : \Omega^\star \to \mathscr{L}(U,U)$ are independent [44, p. 399] and $\mathbb{E}[\|T(\xi)\|_{\mathscr{L}(U,U)}] = \mathbb{E}[\|K(\xi)\|_{\mathscr{L}(U,U)}^2] < \infty$, the strong law of large numbers [159, Thm. 3.3.10] implies $\mathrm{E}^N[T(\xi)] \to \mathbb{E}[T(\xi)]$ as $N \to \infty$ w.p. 1. Therefore, there exists $\omega \in \Omega^\star$ such that $\mathrm{E}^N[T(\xi)](\omega) \to \mathbb{E}[T(\xi)]$ as $N \to \infty$. Since $T(\xi)$ is compact for all $\xi \in \Xi$ and $\mathrm{E}^N[T(\xi)](\omega)$ is the finite sum of compact operators, $\mathrm{E}^N[T(\xi)](\omega)$ is compact [196, p. 407]. Putting together the pieces, we find that $\mathbb{E}[T(\xi)]$ is the uniform limit of the compact operators and, hence, it is compact [196, Thm. 8.1-5]. Furthermore, $\mathbb{E}[T(\xi)]$ is self-adjoint [150, Thm. 3.8.1]. Combined with $\dim(U) = \infty$, we conclude that $\mathbb{E}[T(\xi)]$ cannot have a bounded inverse [196, p. 428] implying that $\mathbb{E}[T(\xi)]$ is not coercive [196, p. 101], [46, Lem. 4.123]. Since $(v, \nabla^2 F_1(u)v)_U = (v, \mathbb{E}[T(\xi)v])_U = (v, \mathbb{E}[T(\xi)]v)_U$ for all $u, v \in U$ (see Lemma 3.2.23 and [150, p. 85]), we conclude that F_1 is not strongly convex.

3.3 Finite Element Discretization and SAA

We discretize an instance of the SAA problem (3.2.2) using finite elements, and derive reliable bounds on the error between its (random) optimal control and the minimizer of the stochastic, linear-quadratic control problem (3.2.27).

We consider the discretized SAA problem

$$\min_{u \in U_{\mathrm{ad},h}} F_{h,N}(u,\omega) + \gamma\|u\|_{L^1(\mathcal{D})}, \tag{3.3.1}$$

where $\gamma \geq 0$, and U_h and Y_h are nonempty, closed finite-dimensional subspaces of $U = L^2(\mathcal{D})$ and $Y = H_0^1(\mathcal{D})$, respectively. The definition of U_h and Y_h is provided in section 3.3.1, and that of $U_{\mathrm{ad},h} \subset U_h$ in (3.3.8). Throughout the section, $\mathcal{D} \subset \mathbb{R}^d$ is a bounded domain.

We define the discretized sample average function $F_{h,N} : U \times \Omega^\star \to \mathbb{R}$ by

$$F_{h,N}(u,\omega) = (1/2)\mathrm{E}^N[\|S_h(u,\xi(\omega)) - y_d\|^2_{L^2(\mathcal{D})}] + (\alpha/2)\|u\|^2_{L^2(\mathcal{D})}, \tag{3.3.2}$$

where $\mathrm{E}^N[\|S_h(u,\xi(\omega)) - y_d\|^2_{L^2(\mathcal{D})}] = (1/N)\sum_{i=1}^N \|S_h(u,\xi^i(\omega)) - y_d\|^2_{L^2(\mathcal{D})}$ and $\xi^i : \Omega^\star \to \Xi$ ($i = 1,\dots,N$) are independent with the same probability distribution as that of ξ. The probability space $(\Omega^\star, \mathcal{F}^\star, P^\star)$ is as in section 3.2. We often drop the second argument of $F_{h,N}$.
The discretized parameterized solution operator $S_h : U \times \Xi \to Y_h$ is defined as follows: for $(u,\xi) \in U \times \Xi$, $S_h(u,\xi) \in Y_h$ is the solution to

$$\text{Find} \quad y_h \in Y_h : \quad (\kappa(\xi)\nabla y_h, \nabla v_h)_{L^2(\mathcal{D})^d} = (u, v_h)_{L^2(\mathcal{D})} \quad \text{for all} \quad v_h \in Y_h. \tag{3.3.3}$$

For simplicity, we assume that $(\kappa(\xi)\nabla y_h, \nabla v_h)_{L^2(\mathcal{D})^d}$ can be evaluated exactly. We consider

$$U_{\mathrm{ad}} = \{\, u \in L^2(\mathcal{D}) : \mathfrak{l} \le u \le \mathfrak{u} \,\}, \quad \mathfrak{l}, \mathfrak{u} \in \mathbb{R}, \quad \mathfrak{l} < 0 < \mathfrak{u}, \tag{3.3.4}$$

which is as in [335, sect. 4.1]. Throughout the section, the inequalities $\mathfrak{l} \le u \le \mathfrak{u}$ are meant in a pointwise almost everywhere sense for $u \in L^2(\mathcal{D})$.
The remaining data for the linear-quadratic control problem (3.2.27) is defined by

$$\begin{aligned} &Y = H^1_0(\mathcal{D}), \quad Z = Y^*, && U = L^2(\mathcal{D}) = H, \\ &Q(\xi)y = y, && y_d \in H, && \Psi = \gamma\|\cdot\|_{L^1(\mathcal{D})}, && \gamma \ge 0, \quad \alpha > 0, \\ &B(\xi)u = -u, && g = 0, && \langle A(\xi)y, v\rangle_{Y^*,Y} = (\kappa(\xi)\nabla y, \nabla v)_{L^2(\mathcal{D})^d}. \end{aligned} \tag{3.3.5}$$

In this case, the discretized SAA problem (3.3.1) can be solved using a semismooth Newton method [303, sect. 4], [321, sect. 9.3]. The random coefficient κ satisfies Assumption 3.3.1 (b).

Assumption 3.3.1. (a) *The domain $\mathcal{D} \subset \mathbb{R}^d$ is bounded, has a Lipschitz boundary, and is convex and polyhedral. Furthermore, we have $d \in \{1,2,3\}$.*
(b) *It holds that $\kappa \in L^\infty(\Xi, C^1(\bar{\mathcal{D}}))$, and there exists $\kappa^*_{\min}$, $\kappa^*_{\max} \in \mathbb{R}$ with $0 < \kappa^*_{\min} \le \kappa(\xi) \le \kappa^*_{\max}$ for all $\xi \in \Xi$.*

Assumption 3.3.1 (a) is as in [3, p. 471]. As shown below in Lemma 3.3.10, Assumption 3.3.1 ensures that the solution $S(\cdot,\xi)$ of (3.2.29) defined by the data in (3.3.5) is an element of $H^2(\mathcal{D})$ for all $\xi \in \Xi$.
We define Friedrichs' constant $C_{\mathcal{D}} > 0$ of the domain $\mathcal{D} \subset \mathbb{R}^d$ by

$$C_{\mathcal{D}} = \sup_{v \in H^1_0(\mathcal{D})\setminus\{0\}} \|v\|_{L^2(\mathcal{D})}/|v|_{H^1}. \tag{3.3.6}$$

If Assumption 3.3.1 (a) is satisfied, then $C_{\mathcal{D}} < \infty$ [151, Thm. 1.13].

Lemma 3.3.2. *Consider the linear-quadratic control problem* (3.2.27) *with the data given by* (3.3.4) *and* (3.3.5). *Let Assumption 3.3.1 hold. Then, the following statements hold true:*
(a) *The mapping A defined in* (3.3.5) *is strongly measurable, and Assumption 3.2.21 holds.*
(b) *For each $\xi \in \Xi$, $A(\xi)$ is self-adjoint.*
(c) *For all $(u,\xi) \in L^2(\mathcal{D}) \times \Xi$,*

$$\|S(u,\xi)\|_{H^1_0(\mathcal{D})} \le (C_{\mathcal{D}}/\kappa^*_{\min})\|u\|_{L^2(\mathcal{D})} \quad \text{and} \quad \|S(u,\xi)\|_{L^2(\mathcal{D})} \le (C^2_{\mathcal{D}}/\kappa^*_{\min})\|u\|_{L^2(\mathcal{D})},$$

where S is defined in (3.2.30) *and $C_{\mathcal{D}} > 0$ in* (3.3.6).

(d) *For all* $(u,\xi) \in L^2(\mathcal{D}) \times \Xi$, *we have* $z(u,\xi) = S(S(u,\xi) - y_d, \xi)$ *and*

$$\|z(u,\xi)\|_{H_0^1(\mathcal{D})} \leq (C_{\mathcal{D}}/\kappa_{\min}^*)((C_{\mathcal{D}}^2/\kappa_{\min}^*)\|u\|_{L^2(\mathcal{D})} + \|y_d\|_{L^2(\mathcal{D})}),$$

where the parameterized adjoint state z *is given by* (3.2.31).

(e) *The function* F *defined in* (3.2.3) *is infinitely many times continuously differentiable.*

(f) *For all* $(u,\xi) \in L^2(\mathcal{D}) \times \Xi$, *we have* $0 \leq \widehat{J}(u,\xi) \leq ((C_{\mathcal{D}}^2/\kappa_{\min}^*)^2 + \alpha/2)\|u\|_{L^2(\mathcal{D})}^2 + \|y_d\|_{L^2(\mathcal{D})}^2$, *where* $\widehat{J}$ *is defined in* (3.2.28).

(g) *The risk-neutral problem* (3.2.27) *has a unique optimal solution, and Assumptions 3.2.1, 3.2.3, 3.2.5 and 3.2.17 hold true.*

Proof. (a) Fix $\xi \in \Xi$. Since $\|\cdot\|_{H_0^1(\mathcal{D})} = |\cdot|_{H^1(\mathcal{D})}$ (see p. viii), Assumption 3.3.1 implies $\langle A(\xi)y, y\rangle_{Y^*,Y} \geq \kappa_{\min}^* \|y\|_{H_0^1(\mathcal{D})}^2$ for all $y \in H_0^1(\mathcal{D})$. Moreover, $A(\xi)$ is linear and bounded. Hence, $A(\xi)$ has a bounded inverse [196, p. 101]. The spaces $Y = H_0^1(\mathcal{D})$ and $U = H = L^2(\mathcal{D})$ are separable Hilbert spaces [1, Thms. 2.15 and 3.5], and $Z = H_0^1(\mathcal{D})^*$ is separable [1, Thm. 1.14]. The strong measurability of A follows from Lemma 3.2.24 and [159, Cors. 1.1.11 and 1.1.24].
(b) The computations in [151, p. 62] imply that $A(\xi)$ is self-adjoint.
(c) The first bound follows from [69, eq. (2.1)], and the second estimate from the first one and (3.3.6).
(d) Using part (b), (3.2.31) and (3.2.29), we obtain $z(u,\xi) = S(S(u,\xi) - y_d, \xi)$. Combined with part (c), we obtain the stability estimate.
(e) Using Lemma 3.2.23, and parts (b) and (c), we deduce the assertions.
(f) Using the definition of $\widehat{J}$ (see (3.2.28)) and Young's inequality, we find that $0 \leq \widehat{J}(u,\xi) \leq \|S(u,\xi)\|_{L^2(\mathcal{D})}^2 + \|y_d\|_{L^2(\mathcal{D})}^2 + (\alpha/2)\|u\|_{L^2(\mathcal{D})}^2$. Combined with part (c), we obtain the estimate.
(g) The feasible set U_{ad} defined in (3.3.4) is nonempty, convex, and closed [316, pp. 116–117]. Since $\mathcal{D}$ is bounded (see Assumption 3.3.1 (a)), Hölder's inequality ensures the (Lipschitz) continuity of $\Psi : U \to \mathbb{R}$ (see (3.3.5)) [1, Thm. 2.8]. Part (e) ensures the continuity of the objective function $F + \Psi$ of (3.2.27). Moreover, it is α-strongly convex with $\alpha > 0$ (see (3.3.5)). Combining the pieces, we obtain the existence of a unique minimizer to (3.2.27) [46, Lem. 2.33]. Using Lemma 3.2.22, parts (d) and (e), and the fact that $\widehat{J}(u,\cdot)$ and $\nabla_u \widehat{J}(u,\cdot)$ are essentially bounded for all $u \in U$, we conclude that Assumptions 3.2.1, 3.2.3, 3.2.5 and 3.2.17 are satisfied. □

3.3.1 State and Control Discretization

We introduce the discretization for the state space $Y = H_0^1(\mathcal{D})$ and for the control space $U = L^2(\mathcal{D})$ defined in (3.3.5). We recall that the domain $\mathcal{D} \subset \mathbb{R}^d$ is bounded.

Assumption 3.3.3. *There exists a sequence* $(Y_h)_{h>0}$ *of nested, closed and finite-dimensional subspaces of* $H_0^1(\mathcal{D})$, *and a constant* $C_Y > 0$, *independent of* $h > 0$, *such that*

$$\inf_{v_h \in Y_h} |v - v_h|_{H^1(\mathcal{D})} \leq C_Y h \|v\|_{H^2(\mathcal{D})} \quad \text{for all} \quad v \in H_0^1(\mathcal{D}) \cap H^2(\mathcal{D}) \quad \text{and} \quad h > 0. \tag{3.3.7}$$

Assumption 3.3.3 is satisfied if Assumption 3.3.1 (a) holds true and Y_h is the space of piecewise linear finite elements on the domain $\mathcal{D}$ [54, sect. 4.4], [3, Lem. 4.3]. The following assumption is based on [335, Assumption 4.1] and [93, Assumption 3.3].

Assumption 3.3.4. *For each* $h > 0$, *there exists* $n_h \in \mathbb{N}$ *and* $\phi_h^j \in L^\infty(\mathcal{D})$ *with* $\phi_h^j \geq 0$ *and* $\|\phi_h^j\|_{L^\infty(\mathcal{D})} = 1$ *for* $j = 1, \ldots, n_h$, *and* $\sum_{j=1}^{n_h} \phi_h^j(x) = 1$ *for almost every* $x \in \mathcal{D}$. *For* $h > 0$, *we define* $U_h = \operatorname{span}\{\, \phi_h^j : j = 1, \ldots, n_h \,\}$. *The sequence* $(U_h)_{h>0}$ *is nested.*

When Assumption 3.3.1 (a) is fulfilled, and either piecewise constant or piecewise linear finite elements are chosen, then Assumption 3.3.4 is fulfilled [93, Rem. 3.1]. Let Assumption 3.3.4 be satisfied and $h > 0$. We define

$$U_{\mathrm{ad},h} = \Big\{ \sum_{j=1}^{n_h} u_j \phi_h^j \in U_h : \quad u_j \in \mathbb{R}, \quad \mathfrak{l} \le u_j \le \mathfrak{u}, \quad j = 1, \dots, n_h \Big\}. \tag{3.3.8}$$

Following [64, Def. 2.2], [93, eqns. (10) and (11)] and [335, p. 868], let us define the quasi-interpolation operator $\mathcal{I}_h : L^1(\mathcal{D}) \to U_h$ by

$$\mathcal{I}_h u = \sum_{j=1}^{n_h} \pi_h^j[u] \phi_h^j, \quad \pi_h^j : L^1(\mathcal{D}) \to \mathbb{R}, \quad \pi_h^j[u] = (\phi_h^j, u)_{L^2(\mathcal{D})} / (\phi_h^j, 1)_{L^2(\mathcal{D})}. \tag{3.3.9}$$

Assumption 3.3.4 and Hölder's inequality [151, Lem. 1.3] imply that $\mathcal{I}_h$ is well-defined.

Assumption 3.3.5. *There exists a constant $C_U > 0$ independent of $h > 0$ such that, for all $h > 0$ and every $u \in H^1(\mathcal{D})$,*

$$\|u - \mathcal{I}_h u\|_{L^2(\mathcal{D})} \le C_U h |u|_{H^1(\mathcal{D})} \quad \textit{and} \quad \|u - \mathcal{I}_h u\|_{H^1(\mathcal{D})^*} \le C_U h^2 \|u\|_{H^1(\mathcal{D})}, \tag{3.3.10}$$

where $\mathcal{I}_h$ is defined in (3.3.9).

According to [93, Lems. 4.3 and 4.4], Assumption 3.3.5 is fulfilled if [93, Assumptions 2.2 and 3.3] hold. We summarize properties of the discretized feasible set $U_{\mathrm{ad},h}$ (see (3.3.8)) and of the interpolation operator $\mathcal{I}_h$ (see (3.3.9)).

Lemma 3.3.6. *If Assumption 3.3.5 holds, $\mathcal{D} \subset \mathbb{R}^d$ is a bounded domain, and $h > 0$, then the following statement hold true:*
(a) *For each $u \in L^1(\mathcal{D})$, we have $\|\mathcal{I}_h u\|_{L^1(\mathcal{D})} \le \|u\|_{L^1(\mathcal{D})}$, where $\mathcal{I}_h$ is defined in* (3.3.9).
(b) *It holds that $U_h \subset L^2(\mathcal{D})$, where U_h is given by Assumption 3.3.5.*
(c) *We have $U_{ad,h} \subset U_{ad}$, where U_{ad} is defined in* (3.3.4) *and $U_{ad,h}$ in* (3.3.8).
(d) *For all $u \in U_{ad}$, we have $\mathcal{I}_h u \in U_{ad,h}$.*

Proof. (a) The statement is shown in [335, p. 870]. Nevertheless, we provide a proof. For all $v \in L^1(\mathcal{D})$, (3.3.9) ensures

$$(v - \pi_h^j[v], \phi_h^i)_{L^2(\mathcal{D})} = (v, \phi_h^j)_{L^2(\mathcal{D})} - \pi_h^j[v](\phi_h^j, 1)_{L^2(\mathcal{D})} = 0 \quad \text{for} \quad j = 1, \dots, n_h, \tag{3.3.11}$$

see also [335, eq. (4.3)], [93, p. 261]. Fix $u \in L^1(\mathcal{D})$. Define $u_+ = \max\{0, u\}$ and $u_- = \min\{0, u\}$. We have u_+, $u_- \in L^1(\mathcal{D})$. Using $\phi_h^j \ge 0$ and $\sum_{j=1}^{n_h} \phi_h^j = 1$ (see Assumption 3.3.4), (3.3.9) and (3.3.11), we obtain $\mathcal{I}_h u_+ \ge 0$ and

$$\|\mathcal{I}_h u_+\|_{L^1(\mathcal{D})} = (\mathcal{I}_h u_+, 1)_{L^2(\mathcal{D})} = \sum_{i=1}^{n_h} (u_+, \phi_h^j)_{L^2(\mathcal{D})} = \|u_+\|_{L^1(\mathcal{D})}.$$

Similary, we can show that $\|\mathcal{I}_h u_-\|_{L^1(\mathcal{D})} = \|u_-\|_{L^1(\mathcal{D})}$. Combined with the linearity of $\mathcal{I}_h$, the triangle inequality, and the definition of the $L^1(\mathcal{D})$-norm, we conclude that

$$\|\mathcal{I}_h u\|_{L^1(\mathcal{D})} \le \|\mathcal{I}_h u_+\|_{L^1(\mathcal{D})} + \|\mathcal{I}_h u_-\|_{L^1(\mathcal{D})} = \|u_+\|_{L^1(\mathcal{D})} + \|u_-\|_{L^1(\mathcal{D})} = \|u\|_{L^1(\mathcal{D})}.$$

(b) Fix $v \in U_h$. Assumption 3.3.4 ensures the existence of $w \in \mathbb{R}^{n_h}$ with $v = \sum_{i=1}^{n_h} w_j \phi_h^j$, and $\phi_h^j \ge 0$ and $\sum_{j=1}^{n_h} \phi_h^j = 1$. Hence $\|v\|_{L^\infty(\mathcal{D})} \le \max_{1 \le j \le n_h} |w_j|$ and $v \in L^2(\mathcal{D})$ [1, Thm. 2.8].

(c) Let $v \in U_{\mathrm{ad},h}$ be arbitrary. Using (3.3.8), we deduce the existence of $w_j \in \mathbb{R}$ with $v = \sum_{j=1}^{n_h} w_j \phi_h^j$ and $\mathfrak{l} \leq w_j \leq \mathfrak{u}$. Since $\phi_h^j \geq 0$ (see Assumption 3.3.4), we obtain $\phi_h^j \mathfrak{l} \leq \phi_h^j u_j \leq \phi_h^j \mathfrak{u}$ for $j = 1, \ldots, n_h$. Combined with $\sum_{j=1}^{n_h} \phi_h^j = 1$, we deduce $\mathfrak{l} \leq v \leq \mathfrak{u}$. Hence, $v \in U_{\mathrm{ad}}$.
(d) Fix $u \in U_{\mathrm{ad}}$ and $j \in \{1, \ldots, n_h\}$. Since Assumption 3.3.4 holds, we have $\phi_h^j \geq 0$. Combined with (3.3.4) and (3.3.9), we find that $\mathfrak{l} \leq \pi_h^j[u] \leq \mathfrak{u}$. Now, the definition of $U_{\mathrm{ad},h}$ provided in (3.3.8) and $u \in L^1(\mathcal{D})$ [1, Thm. 2.8] ensure $\mathcal{I}_h u \in U_{\mathrm{ad},h}$. □

Lemma 3.3.7. *Consider the discretized SAA problem* (3.3.1) *with the data given by* (3.3.4) *and* (3.3.5). *Suppose that Assumptions 3.3.1 and 3.3.3–3.3.5 are fulfilled. Then, the following statements hold true:*

(a) *For each $\omega \in \Omega^\star$, the function $F_{h,N}(\cdot, \omega)$ defined in* (3.3.2) *is infinitely many times continuously differentiable.*
(b) *For each $\omega \in \Omega^\star$, the discretized SAA problem* (3.3.1) *has a unique optimal solution $u_{h,N}^*(\omega)$, and $u_{h,N}^* : \Omega^\star \to U_h$ is measurable.*

Proof. (a) The spaces Y_h and U_h are Banach spaces according to Assumptions 3.3.3 and 3.3.4 and Lemma 3.3.6. Combined with Lemmas 3.2.23 and 3.3.2, we deduce the assertion.
(b) Fix $\omega \in \Omega^\star$. Part (a) ensures the continuity of $F_{h,N}(\cdot, \omega)$. Moreover, it is α-strongly convex with $\alpha > 0$ (see (3.3.5)). Furthermore, the feasible set U_{ad} defined in (3.3.4) is nonempty, closed, convex (and bounded) (see Lemma 3.3.2). Putting together the pieces, we find that (3.3.1) has a unique optimal solution $u_{h,N}^*(\omega)$ [46, Lem. 2.33]. Since U_h is a Banach space, Lemma 3.2.4 ensures the measurability of $u_{h,N}^* : \Omega^\star \to U_h$. □

3.3.2 Reliable Error Estimates

We derive reliable bounds on the $L^2(\mathcal{D})$-distance between the optimal solution of the discretized SAA problem (3.3.1) and the minimizer of the risk-neutral problem (3.2.27).
Throughout the section, the linear-quadratic control problem (3.2.27) is considered with the data given by (3.3.4) and (3.3.5).

Proposition 3.3.8. *Suppose that Assumptions 3.3.1 and 3.3.3–3.3.5 are fulfilled. Let u^* be the minimizer of the risk-neutral problem* (3.2.27) *and for each $\omega \in \Omega^\star$, let $u_{h,N}^*(\omega)$ be the optimal solution of the discretized SAA problem* (3.3.1). *Let $\varepsilon > 0$, $\delta \in (0,1)$, and $h \in (0,1)$ be arbitrary. If $N \geq \ln(2/\delta)/\varepsilon^2$, then with a probability of at least $1 - \delta$,*

$$\|u_{h,N}^* - u^*\|_{L^2(\mathcal{D})} \leq c\big(\tfrac{h+\varepsilon}{\alpha}\big)\big(\|u^*\|_{L^2(\mathcal{D})} + \|y_d\|_{L^2(\mathcal{D})}\big) + ch\big(1 + \tfrac{1}{\alpha}\big)\|u^*\|_{H^1(\mathcal{D})}. \tag{3.3.12}$$

Here, $c > 0$ is a deterministic constant that is independent of $h > 0$, $\alpha > 0$, $\varepsilon > 0$ and $\delta \in (0,1)$, but depends on $C_{\mathcal{D}} > 0$ (see (3.3.6)*), $\kappa_{\min}^*$, $\kappa_{\max}^*$, $\|\kappa\|_{L^\infty(\Xi; C^1(\bar{\mathcal{D}}))}$ (see Assumption 3.3.1), $C_{H^2} > 0$ (see Lemma 3.3.10), $C_Y > 0$ (see Assumption 3.3.3), and $C_U > 0$ (see Assumption 3.3.5).*

Proposition 3.3.8 implies that random control $u_{h,N}^*$ is contained in an $L^2(\mathcal{D})$-ball about the minimizer u^* of the true problem (3.2.27) with high probability.
A bound on $\|u^*\|_{H^1(\mathcal{D})}$ is provided in Lemma 3.3.9. We obtain $\|u^*\|_{L^2(\mathcal{D})} \leq (|\mathfrak{l}| + \mathfrak{u})\|1\|_{L^2(\mathcal{D})}$ using the definition of U_{ad} (see (3.3.4)). The constant $c > 0$ in Proposition 3.3.8 may be difficult to estimate as it depends, for example, on those provided by Assumptions 3.3.3 and 3.3.5. The term $(1/\alpha)\|u^*\|_{H^1(\mathcal{D})}$ in (3.3.12) might not be optimal; cf. [335, Prop. 4.5].
We prove Proposition 3.3.8 using Lemmas 3.3.7 and 3.3.9–3.3.11.

Lemma 3.3.9. *If Assumption 3.3.1 holds, U_{ad} is given by* (3.3.4), *then $\mathbb{E}[z(u^*, \xi)] \in H_0^1(\mathcal{D})$, $u^* \in H^1(\mathcal{D})$ and $\nabla F(u^*) \in H^1(\mathcal{D})$, where F is defined in* (3.2.3) *and z in* (3.2.31). *Moreover, $\|u^*\|_{H^1(\mathcal{D})} \leq (1/\alpha)\|\mathbb{E}[z(u^*, \xi)]\|_{H^1(\mathcal{D})} + (|\mathfrak{l}| + \mathfrak{u})\|1\|_{L^2(\mathcal{D})}$.*

Proof. Lemma 3.2.22 reveals the (strong) measurability of the parameterized adjoint state $z(u^*,\cdot) : \Xi \to H_0^1(\mathcal{D})$ defined in (3.2.31), and Lemma 3.3.2 ensures $\mathbb{E}[\|z(u^*,\xi)\|_{H_0^1(\mathcal{D})}] < \infty$. Consequently, $z(u^*,\xi)$ is Bochner integrable. We obtain $\mathbb{E}[z(u^*,\xi)] \in H_0^1(\mathcal{D})$ [159, p. 14]. Since $\Psi = \gamma\|\cdot\|_{L^1(\mathcal{D})}$ (see (3.3.5)) is convex and continuous, the necessary and sufficient optimality conditions for (3.2.1) can be expressed as the deterministic variational inequality $(\mathsf{p} + \alpha u^* + \lambda^*, u - u^*)_{L^2(\mathcal{D})} \geq 0$ for all $u \in U_{\mathrm{ad}}$, where $\mathsf{p} = \mathbb{E}[z(u^*,\xi)]$ and λ^* is the Riesz representation of an element of $\partial\Psi(u^*)$ [163, sect. 3], [46, Chap. 2]. Since U_{ad} is given by (3.3.4) and $U = L^2(\mathcal{D})$ (see (3.3.5)), this variational inequality allows for a pointwise characterization [303, sect. 2 and eq. (4.4)], [334, eq. (2)], [265, pp. 94–95]. Combining this pointwise characterization with [178, Cor. A.5 on p. 54], we obtain $u^* \in H^1(\mathcal{D})$. Now, Lemma 3.2.23 and (3.2.32) ensure $\nabla F(u^*) \in H^1(\mathcal{D})$. The bound on $\|u^*\|_{H^1(\mathcal{D})}$ follows from that in [335, p. 870], when combined with the fact that the above variational inequality is deterministic and that $\mathrm{D}\mathfrak{l} = \mathrm{D}\mathfrak{u} = 0$. □

We recall that relations between random variables hold w.p. 1 if not stated otherwise (see p. viii).

Lemma 3.3.10. *Let Assumptions 3.3.1 and 3.3.3–3.3.5 hold, and let $h > 0$. Then, the following statements hold true:*

(a) *There exists $C_{H^2} > 0$ such that $\|S(u,\xi)\|_{H^2(\mathcal{D})} \leq C_{H^2}C_1\|u\|_U$ and $S(u,\xi) \in H^2(\mathcal{D})$ for all $(u,\xi) \in U \times \Xi$, where S is defined in* (3.2.30)*, and $C_1 = (\kappa_{\max}^*/(\kappa_{\min}^*)^4)\|\kappa\|^2_{L^\infty(\Xi;C^1(\bar{\mathcal{D}}))}$.*

(b) *$|S(u,\xi) - S_h(u,\xi)|_{H^1(\mathcal{D})} \leq C_Y C_{H^2} C_2 h\|u\|_U$ for all $(u,\xi) \in U \times \Xi$. Here, $C_Y > 0$ is defined by Assumption 3.3.3 and $S_h : L^2(\mathcal{D}) \times \Xi \to Y_h$ in* (3.3.3)*, and $C_2 = C_1(\kappa_{\max}^*/\kappa_{\min}^*)^{1/2}$.*

(c) *$\|S(u,\xi) - S_h(u,\xi)\|_{L^2(\mathcal{D})} \leq \kappa_{\max}^* C_Y^2 C_{H^2}^2 C_2 h^2\|u\|_{L^2(\mathcal{D})}$ for all $(u,\xi) \in U \times \Xi$.*

(d) *For $C_{\mathcal{D}} > 0$ defined in* (3.3.6)*, $F_{h,N}$ in* (3.3.2) *and $\mathcal{I}_h$ in* (3.3.9)*, we have*

$$|(\nabla F_{h,N}(\mathcal{I}_h u^*) - \nabla F_{h,N}(u^*), \mathcal{I}_h u^* - u_{h,N}^*)_U| \leq \Big(\alpha + \tfrac{C_{\mathcal{D}}^4}{(\kappa_{\min}^*)^2}\Big)\|\mathcal{I}_h u^* - u_{h,N}^*\|_U\|\mathcal{I}_h u^* - u^*\|_U.$$

(e) *$\|\nabla F_{h,N}(u^*) - \nabla F_N(u^*)\|_U \leq C_3 h^2(\|u^*\|_U + \|y_d\|_U)$, where the constant $C_3 > 0$ is defined by $C_3 = 2\max\{(C_{\mathcal{D}}^2/\kappa_{\min}^*), 1\}\kappa_{\max}^* C_Y^2 C_{H^2}^2 C_2$.*

(f) *$|(\nabla F(u^*), \mathcal{I}_h u^* - u^*)_{L^2(\mathcal{D})}| \leq C_U h^2\|\nabla F(u^*)\|_{H^1(\mathcal{D})}\|u^*\|_{H^1(\mathcal{D})}$, where $C_U > 0$ is provided by Assumption 3.3.5, and F is defined in* (3.2.3)*.*

Proof. (a) See [3, Thm. 3.1].
(b) The assertion follows essentially from the proof of [70, Thm. 3.9]. Using Assumption 3.3.1 and Céa's lemma [70, Lem. 3.8], we find that

$$|S(u,\xi) - S_h(u,\xi)|_{H^1(\mathcal{D})} \leq (\kappa_{\max}^*/\kappa_{\min}^*)^{1/2} \inf_{v_h \in Y_h} |S(u,\xi) - v_h|_{H^1(\mathcal{D})}.$$

Combined with Assumption 3.3.3 and part (a), we obtain

$$|S(u,\xi) - S_h(u,\xi)|_{H^1(\mathcal{D})} \leq (\kappa_{\max}^*/\kappa_{\min}^*)^{1/2} C_Y h\|S(u,\xi)\|_{H^2(\mathcal{D})} \leq (\kappa_{\max}^*/\kappa_{\min}^*)^{1/2} C_Y C_{H^2} C_1 h\|u\|_U,$$

where $C_1 > 0$ is defined in part (a).
(c) The assertion follows from the proof of [3, Thm. 4.4]. Tracing the constants in the proof of [3, Thm. 4.4], we obtain the estimate.
(d) Lemmas 3.2.23 and 3.3.7 and the fact that $F_{h,N}$ is quadratic (see (3.3.2)) yield

$$\begin{aligned}(\nabla F_{h,N}(\mathcal{I}_h u^*) - \nabla F_{h,N}(u^*), \mathcal{I}_h u^* - u_{h,N}^*)_U &= \mathrm{E}^N[(S_h(\mathcal{I}_h u^* - u^*, \xi), S_h(\mathcal{I}_h u^* - u_{h,N}^*, \xi))_U] \\ &\quad + \alpha(\mathcal{I}_h u^* - u^*, \mathcal{I}_h u^* - u_{h,N}^*)_U.\end{aligned}$$

Using similar arguments as in Lemma 3.3.2, we obtain $\|S_h(u,\xi)\|_{L^2(\mathcal{D})} \leq (C_{\mathcal{D}}^2/\kappa_{\min}^*)\|u\|_{L^2(\mathcal{D})}$ for all $u \in U$. Combined with the Cauchy–Schwarz inequality, we obtain the assertion.

(e) Using Lemmas 3.2.23 and 3.3.2, we find that

$$\nabla F_{h,N}(u^*) - \nabla F_N(u^*) = \mathrm{E}^N[S_h(S_h(u^*,\xi) - y_d, \xi)] - \mathrm{E}^N[S(S(u^*,\xi) - y_d, \xi)],$$

where E^N is the sample mean (see p. viii). We separately estimate $S_h(S(u^*,\xi) - y_d, \xi) - S(S(u^*,\xi) - y_d, \xi)$ and $S_h(S_h(u^*,\xi) - y_d, \xi) - S_h(S(u^*,\xi) - y_d, \xi)$ for $\xi \in \Xi$. Using part (c) and Lemma 3.3.2, we find that

$$\begin{aligned}\|S_h(S_h(u^*,\xi) - y_d, \xi) - S_h(S(u^*,\xi) - y_d, \xi)\|_{L^2(\mathcal{D})} &\leq (C_{\mathcal{D}}^2/\kappa_{\min}^*)\|S_h(u^*,\xi) - S(u^*,\xi)\|_{L^2(\mathcal{D})} \\ &\leq (C_{\mathcal{D}}^2/\kappa_{\min}^*)\kappa_{\max}^* C_Y^2 C_{H^2}^2 C_2 h^2 \|u^*\|_{L^2(\mathcal{D})}.\end{aligned}$$

Using part (c) and Lemma 3.3.2, we further find that

$$\|S_h(S(u^*,\xi) - y_d, \xi) - S(S(u^*,\xi) - y_d, \xi)\|_{L^2(\mathcal{D})} \leq \kappa_{\max}^* C_Y^2 C_{H^2}^2 C_2 h^2 \|S(u^*,\xi) - y_d\|_{L^2(\mathcal{D})}.$$

The triangle inequality and Lemma 3.3.2 also yield $\|S(u^*,\xi) - y_d\|_{L^2(\mathcal{D})} \leq (C_{\mathcal{D}}^2/\kappa_{\min}^*)\|u^*\|_{L^2(\mathcal{D})} + \|y_d\|_{L^2(\mathcal{D})}$. Putting together the pieces, we obtain the assertion.
(f) Since $H^1(\mathcal{D}) \hookrightarrow L^2(\mathcal{D}) \hookrightarrow H^1(\mathcal{D})^*$ is a Gelfand triple [316, p. 147], the embedding $L^2(\mathcal{D}) \hookrightarrow H^1(\mathcal{D})^*$ is given by $\langle v, w\rangle_{H^1(\mathcal{D})^*, H^1(\mathcal{D})} = (v,w)_{L^2(\mathcal{D})}$ for all $v \in L^2(\mathcal{D})$ and $w \in H^1(\mathcal{D})$ [151, Rem. 1.17]. Combined with $u^* \in H^1(\mathcal{D})$, $\nabla F(u^*) \in H^1(\mathcal{D})$ (see Lemma 3.3.9), and $\mathcal{I}_h u^* \in L^2(\mathcal{D})$ (see Lemma 3.3.6), we have $|(\nabla F(u^*), \mathcal{I}_h u^* - u^*)_{L^2(\mathcal{D})}| \leq \|\nabla F(u^*)\|_{H^1(\mathcal{D})} \|\mathcal{I}_h u^* - u^*\|_{H^1(\mathcal{D})^*}$. Together with Assumption 3.3.5, we obtain the assertion. □

Lemma 3.3.11. *If Assumptions 3.3.1 and 3.3.3–3.3.5 hold, and $h > 0$, then w.p. 1,*

$$\begin{aligned}\alpha\|\mathcal{I}_h u^* - u_{h,N}^*\|_U^2 \leq &(\nabla F_{h,N}(\mathcal{I}_h u^*) - \nabla F_{h,N}(u^*), \mathcal{I}_h u^* - u_{h,N}^*)_U \\ &+ (\nabla F_{h,N}(u^*) - \nabla F_N(u^*), \mathcal{I}_h u^* - u_{h,N}^*)_U \\ &+ (\nabla F_N(u^*) - \nabla F(u^*), \mathcal{I}_h u^* - u_{h,N}^*)_U \\ &+ (\nabla F(u^*), \mathcal{I}_h u^* - u^*)_U,\end{aligned} \tag{3.3.13}$$

where $\mathcal{I}_h$ is defined in (3.3.9)*, $F_{h,N}$ in* (3.3.2)*, and F and F_N in* (3.2.3)*. Here, for each $\omega \in \Omega^\star$, $u_{h,N}^*(\omega)$ is the optimal solution of* (3.3.1) *and u^* is that of* (3.2.1)*.*

Proof. The proof is inspired by the arguments used by Meidner and Vexler [231, Thm. 5.2]. Lemma 3.3.6 and $\Psi = \gamma\|\cdot\|_{L^1(\mathcal{D})}$ with $\gamma \geq 0$ (see (3.3.5)) ensure $u_{h,N}^* \in U_{\mathrm{ad}}$, $\mathcal{I}_h u^* \in U_{\mathrm{ad},h}$, and $\Psi(\mathcal{I}_h u^*) \leq \Psi(u^*)$. Lemmas 3.3.2 and 3.3.7 imply that F and $F_{h,N}$ are continuously differentiable and convex. Using $u_{h,N}^* \in U_{\mathrm{ad}}$ and Lemma 3.2.14, we find that $(\nabla F(u^*), u_{h,N}^* - u^*)_U + \Psi(u_{h,N}^*) - \Psi(u^*) \geq 0$. Using the same arguments as in the proof of Lemma 3.2.14 and $\mathcal{I}_h u^* \in U_{\mathrm{ad},h}$, we obtain $(\nabla F_{h,N}(u_{h,N}^*), \mathcal{I}_h u^* - u_{h,N}^*)_U + \Psi(\mathcal{I}_h u^*) - \Psi(u_{h,N}^*) \geq 0$. Adding these two inequalities and using $\Psi(\mathcal{I}_h u^*) \leq \Psi(u^*)$ yields

$$0 \leq (\nabla F(u^*), u_{h,N}^* - u^*)_U + (\nabla F_{h,N}(u_{h,N}^*), \mathcal{I}_h u^* - u_{h,N}^*)_U. \tag{3.3.14}$$

Since $F_{h,N}$ is α-strongly convex (see (3.3.2)) and Gâteaux differentiable, we have

$$\alpha\|\mathcal{I}_h u^* - u_{h,N}^*\|_U^2 \leq (\nabla F_{h,N}(\mathcal{I}_h u^*) - \nabla F_{h,N}(u_{h,N}^*), \mathcal{I}_h u^* - u_{h,N}^*)_U.$$

Adding this inequality and the estimate (3.3.14), we conclude that

$$\begin{aligned}\alpha\|\mathcal{I}_h u^* - u_{h,N}^*\|_U^2 &\leq (\nabla F_{h,N}(\mathcal{I}_h u^*), \mathcal{I}_h u^* - u_{h,N}^*)_U + (\nabla F(u^*), u_{h,N}^* - u^*)_U \\ &= (\nabla F_{h,N}(\mathcal{I}_h u^*), \mathcal{I}_h u^* - u_{h,N}^*)_U + (\nabla F(u^*), \mathcal{I}_h u^* - u^*)_U \\ &\quad + (\nabla F(u^*), u_{h,N}^* - \mathcal{I}_h u^*)_U.\end{aligned}$$

Manipulating the term on the right-hand side and using the measurability of $u_{h,N}^* : \Omega^\star \to U_h$ (see Lemma 3.3.7), we obtain (3.3.13). □

Proof of Proposition 3.3.8. The triangle inequality and Assumption 3.3.5 yield

$$\|u^*_{h,N} - u^*\|_U \leq C_U h |u^*|_{H^1(\mathcal{D})} + \|u^*_{h,N} - \mathcal{I}_h u^*\|_U, \tag{3.3.15}$$

where $\mathcal{I}_h$ is defined in (3.3.9) and $C_U > 0$ is given by Assumption 3.3.5. Lemma 3.3.7 ensures the measurability of $u^*_{h,N} : \Omega^\star \to U_h$.
In the following steps, we derive a bound on $\|u^*_{h,N} - \mathcal{I}_h u^*\|_U$ using Lemmas 3.3.2, 3.3.10 and 3.3.11 and Theorem 3.2.16.
Using the Cauchy–Schwarz inequality, $(2\rho_1\rho_2)^{1/2} \leq \rho_1 + \rho_2$, valid for all ρ_1, $\rho_2 \in \mathbb{R}_+$ and Lemma 3.3.10, we find that

$$\begin{aligned}
|2(\nabla F_{h,N}(\mathcal{I}_h u^*) - \nabla F_{h,N}(u^*), \mathcal{I}_h u^* - u^*_{h,N})_U|^{1/2} &\leq (4/\alpha^{1/2})(\alpha + C^4_{\mathcal{D}}/(\kappa^*_{\min})^2)\|\mathcal{I}_h u^* - u^*\|_{L^2(\mathcal{D})} \\
&\quad + (\alpha^{1/2}/4)\|\mathcal{I}_h u^* - u^*_{h,N}\|_{L^2(\mathcal{D})}, \\
|2(\nabla F_{h,N}(u^*) - \nabla F_N(u^*), \mathcal{I}_h u^* - u^*_{h,N})_U|^{1/2} &\leq (4/\alpha^{1/2})\|\nabla F_{h,N}(u^*) - \nabla F_N(u^*)\|_U \\
&\quad + (\alpha^{1/2}/4)\|\mathcal{I}_h u^* - u^*_{h,N}\|_U, \\
|2(\nabla F_N(u^*) - \nabla F(u^*), \mathcal{I}_h u^* - u^*_{h,N})_U|^{1/2} &\leq (1/\alpha^{1/2})\|\nabla F_N(u^*) - \nabla F(u^*)\|_U \\
&\quad + (\alpha^{1/2}/4)\|\mathcal{I}_h u^* - u^*_{h,N}\|_U,
\end{aligned}$$

where $C_{\mathcal{D}} > 0$ is defined in (3.3.6). Combined with (3.3.13) and $(\rho_1 + \rho_2)^{1/2} \leq \rho_1^{1/2} + \rho_2^{1/2}$ valid for all ρ_1, $\rho_2 \in \mathbb{R}_+$, we conclude that

$$\begin{aligned}
(1/8)\|u^*_{h,N} - \mathcal{I}_h u^*\|_{L^2(\mathcal{D})} \leq\ & (1/\alpha)(\alpha + C^4_{\mathcal{D}}/(\kappa^*_{\min})^2)\|\mathcal{I}_h u^* - u^*\|_{L^2(\mathcal{D})} \\
&+ (1/\alpha)\|\nabla F_{h,N}(u^*) - \nabla F_N(u^*)\|_{L^2(\mathcal{D})} \\
&+ (1/\alpha)\|\nabla F_N(u^*) - \nabla F(u^*)\|_{L^2(\mathcal{D})} \\
&+ (1/\alpha^{1/2})|(\nabla F(u^*), \mathcal{I}_h u^* - u^*)_{L^2(\mathcal{D})}|^{1/2}.
\end{aligned}$$

Combined with Lemma 3.3.10 and Assumption 3.3.5, we further find that

$$\begin{aligned}
(1/8)\|u^*_{h,N} - \mathcal{I}_h u^*\|_{L^2(\mathcal{D})} \leq\ & (1/\alpha)(\alpha + C^4_{\mathcal{D}}/(\kappa^*_{\min})^2) C_U h |u^*|_{H^1(\mathcal{D})} \\
&+ (1/\alpha) C_3 h^2 (\|u^*\|_{L^2(\mathcal{D})} + \|y_d\|_U) \\
&+ (1/\alpha)\|\nabla F_N(u^*) - \nabla F(u^*)\|_{L^2(\mathcal{D})} \\
&+ (1/\alpha^{1/2}) h (C_U \|\nabla F(u^*)\|_{H^1(\mathcal{D})} \|u^*\|_{H^1(\mathcal{D})})^{1/2},
\end{aligned} \tag{3.3.16}$$

where $C_3 > 0$ is defined in Lemma 3.3.10.
We must yet derive bounds on the third and fourth term in the right-hand side of (3.3.16). The triangle inequality, Jensen's inequality, and Lemmas 3.2.23, 3.3.2 and 3.3.9 imply

$$\begin{aligned}
\|\nabla F(u^*)\|_{H^1(\mathcal{D})} &\leq \alpha \|u^*\|_{H^1(\mathcal{D})} + \|\mathbb{E}[z(u^*, \xi)]\|_{H^1(\mathcal{D})} \\
&\leq \alpha \|u^*\|_{H^1(\mathcal{D})} + (C_{\mathcal{D}}/\kappa^*_{\min})((C^2_{\mathcal{D}}/\kappa^*_{\min})\|u^*\|_{L^2(\mathcal{D})} + \|y_d\|_{L^2(\mathcal{D})}).
\end{aligned}$$

Combined with Young's inequality, we obtain

$$\tfrac{1}{\alpha^{1/2}}\|\nabla F(u^*)\|^{1/2}_{H^1(\mathcal{D})}\|u^*\|^{1/2}_{H^1(\mathcal{D})} \leq 2\|u^*\|_{H^1(\mathcal{D})} + \tfrac{C_{\mathcal{D}}}{\alpha\kappa^*_{\min}}\big(\tfrac{C^2_{\mathcal{D}}}{\kappa^*_{\min}}\|u^*\|_{L^2(\mathcal{D})} + \|y_d\|_{L^2(\mathcal{D})}\big). \tag{3.3.17}$$

Since $N \geq \ln(2/\delta)/\varepsilon^2$, Theorem 3.2.16 ensures[2], with a probability of at least $1 - \delta$,

$$\|\nabla F_N(u^*) - \nabla F(u^*)\|_U \leq 2^{1/2}\varepsilon \|\nabla_u \widehat{J}(u^*, \xi) - \nabla F(u^*)\|_{L^\infty(\Xi; L^2(\mathcal{D}))}. \tag{3.3.18}$$

[2]We define $\tau = \|\nabla_u \widehat{J}(u^*, \xi) - \nabla F(u^*)\|_{L^\infty(\Xi; L^2(\mathcal{D}))} \in \mathbb{R}_{++}$. Theorem 3.2.16 gives $\mathrm{Prob}(\|\nabla F_N(u^*) - \nabla F(u^*)\|_U \geq r) \leq 2\exp(-\tau^{-2} r^2 N/2)$ for all $r > 0$, which is equivalent to $\mathrm{Prob}(\|\nabla F_N(u^*) - \nabla F(u^*)\|_U \geq \sqrt{2}\tau r) \leq 2\exp(-r^2 N)$ for all $r > 0$.

It remains to derive a bound on $\|\nabla_u \widehat{J}(u^*,\xi) - \nabla F(u^*)\|_{L^\infty(\Xi;L^2(\mathcal{D}))}$. Using the definition of the adjoint state z (see (3.2.31)), (3.2.32), and Lemmas 3.2.23 and 3.3.2, we have

$$\begin{aligned}\|\nabla_u \widehat{J}(u^*,\xi) - \nabla F(u^*)\|_{L^\infty(\Xi;L^2(\mathcal{D}))} &= \|z(u^*,\xi) - \mathbb{E}[z(u^*,\xi)]\|_{L^\infty(\Xi;L^2(\mathcal{D}))} \\ &\le 2\|z(u^*,\xi)\|_{L^\infty(\Xi;L^2(\mathcal{D}))} \\ &\le \tfrac{2C_{\mathcal{D}}}{\kappa^*_{\min}}\big(\tfrac{C^2_{\mathcal{D}}}{\kappa^*_{\min}}\|u^*\|_{L^2(\mathcal{D})} + \|y_d\|_{L^2(\mathcal{D})}\big).\end{aligned} \tag{3.3.19}$$

Combining $h \in (0,1)$, $|\cdot|_{H^1(\mathcal{D})} \le \|\cdot\|_{H^1(\mathcal{D})}$ (see p. viii) with (3.3.15), (3.3.16), (3.3.17), (3.3.18), and (3.3.19), we obtain (3.3.12). □

3.4 Risk-Averse Optimization using the Superquantile

We consider risk-averse convex optimization using the superquantile/conditional value-at-risk, and analyze the expected value of the corresponding SAA problem's optimal value. Risk-averse PDE-constrained optimization using the superquantile is considered, for example, in [122, 190, 191, 192, 187, 195, 193, 194].

For $\beta \in (0,1)$, we consider the risk-averse optimization problem

$$\min_{u\in U_{\mathrm{ad}}} \{\, \mathcal{Q}_\beta(\mathcal{J}(u)) + \Psi(u) \,\}, \tag{3.4.1}$$

where U_{ad} and Ψ satisfy Assumption 3.2.1 (a) and Assumption 3.2.1 (c), respectively. For $\beta \in [0,1)$, the *β-superquantile* $\mathcal{Q}_\beta : L^1(\Omega) \to \mathbb{R}$ is defined by

$$\mathcal{Q}_\beta(Z) = \inf_{t\in\mathbb{R}} \big\{\, t + \tfrac{1}{1-\beta}\mathbb{E}[(Z-t)_+] \,\big\}, \tag{3.4.2}$$

see [272, 273]. Here, $(\cdot)_+ = \max\{\cdot, 0\}$.

Moreover, we define $\mathcal{J} : V \to L^1(\Omega)$ by $\mathcal{J}(u)(\omega) = \widehat{J}(u,\xi(\omega))$. Here, $V \subset U$ is a convex neighborhood of U_{ad}. If $\Omega \ni \omega \mapsto \widehat{J}(u,\xi(\omega))$ is integrable for all $u \in V$, and Assumptions 3.2.1 (a) and 3.2.1 (d) are satisfied, then $\mathcal{J}$ is well-defined and convex [161, pp. 9–10]. The function $\mathcal{Q}_\beta \circ \mathcal{J}$ is called a composite risk function in the literature [279, sect. 3.2].

Other names for the superquantile frequently used in the literature are *conditional value-at-risk* [273] and *average value-at-risk* [294, sect. 6.2.4]. We prefer the "application-independent" term superquantile suggested by Rockafellar and Royset [271, p. 503].

We report some properties of the β-superquantile. The β-superquantile is a *coherent risk measure*, that is, $\mathcal{Q}_\beta$ is convex, nondecreasing, translation equivariant and positive homogeneous [269, Thm. 2], [294, pp. 279 and 291]. Furthermore, the β-superquantile is *risk-averse*, that is, $\mathcal{Q}_\beta(Z) > \mathbb{E}[Z]$ for all nondegenerate $Z \in L^1(\Omega)$ and $\beta \in (0,1)$ [27, Thm. 4]. For $Z \in L^\infty(\Omega)$, it holds that $\mathcal{Q}_\beta(Z) \to \|Z\|_{L^\infty(\Omega)}$ as $\beta \to 1$ [294, Rem. 24]. For $\beta = 0$, we have $\mathcal{Q}_\beta(Z) = \mathbb{E}[Z]$ for all $Z \in L^1(\Omega)$ [294, Rem. 24]. Moreover, the function $[0,1) \ni \beta \mapsto \mathcal{Q}_\beta(Z)$ is continuous for each fixed $Z \in L^1(\Omega)$ [294, Rem. 24].

By [273, Thm. 14], problem (3.4.1) is equivalent to

$$\min_{(u,t)\in U_{\mathrm{ad}}\times\mathbb{R}} \big\{\, f(u,t) = t + \tfrac{1}{1-\beta}\mathbb{E}[(\widehat{J}(u,\xi) - t)_+] + \Psi(u) \,\big\}. \tag{3.4.3}$$

Let Assumption 3.2.1 (d) hold. We define $F : U \times \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ and $F_N : U \times \mathbb{R} \times \Omega^\star \to \mathbb{R}$ by

$$F(u,t) = t + \tfrac{1}{1-\beta}\mathbb{E}[(\widehat{J}(u,\xi) - t)_+] \quad \text{and} \quad F_N(u,t,\omega) = t + \tfrac{1}{1-\beta}\mathrm{E}^N[(\widehat{J}(u,\xi(\omega)) - t)_+], \tag{3.4.4}$$

where $\mathrm{E}^N[(\widehat{J}(u,\xi(\omega)) - t)_+] = \sum_{i=1}^N (\widehat{J}(u,\xi^i(\omega)) - t)_+$, and $\xi^i : \Omega^\star \to \Xi$ are independent with the same distribution as that of $\xi : \Omega \to \Xi$. The probability space $(\Omega^\star, \mathcal{F}^\star, P^\star)$ is as in section 3.2.

If $u \in U$, $\widehat{J}(u,\xi) \in L^1(\Omega)$ and $\beta \in (0,1)$, then $(s)_+ \geq s$ and $(s)_+ \geq 0$, valid for all $s \in \mathbb{R}$, ensure that $F(u,\cdot)$ is coercive.
We consider the SAA problem corresponding to (3.4.4)

$$\min_{(u,t)\in U_{\mathrm{ad}}\times T_{\mathrm{ad}}} \big\{\, f_N(u,t,\omega) = F_N(u,t,\omega) + \Psi(u) \,\big\}, \tag{3.4.5}$$

where $T_{\mathrm{ad}} \subset \mathbb{R}$ is a nonempty, compact interval. We often drop the third argument of f_N and of F_N. If (u^*,t^*) is an optimal solution of (3.4.3), we require that $t^* \in T_{\mathrm{ad}}$. Below, we discuss how T_{ad} can be constructed.
When augmenting T_{ad} as feasible set for the variable t in (3.4.3), and U_{ad} is bounded, robust stochastic approximation can be applied to it; see Lan, Nemirovski, and Shapiro [208, sect. 4.2]. For analyzing stochastic programs with superquantile constraints, Wang and Ahmed [339] also have made use of the fact that set $\arg\min_{t\in\mathbb{R}} F(u^*,t)$ can often be bounded effectively, where u^* is an optimal solution to (3.4.1).
We discuss the construction of the feasible set T_{ad} (see (3.4.5)) for a particular example. We follow the approach outlined in [208, p. 441]. Let $\widehat{J}$ be the tracking-type functional defined by (3.2.28) with the data given by (3.3.4) and (3.3.5). Moreover, let $\beta \in (0,1)$. We show that $T_{\mathrm{ad}} = [0, (1/2)\|y_d\|_H^2]$ is a possible choice, where $y_d \in H$ (see (3.3.5)). Let (u^*,t^*) be an optimal solution of (3.4.3). By (3.2.28), we have $\widehat{J} \geq 0$. Hence, (3.4.2) ensures $t^* \geq 0$ (see also Lemma 3.4.5).[3] Using (3.4.2), $(\cdot)_+ \geq 0$, $0 \in U_{\mathrm{ad}}$ (see (3.3.4)), $\Psi \geq 0 = \Psi(0)$ (see (3.3.5)), and the fact that the optimal value of (3.4.1) equals that of (3.4.3) [273, Thm. 14], we find that

$$t^* \leq \mathcal{Q}_\beta(\widehat{J}(u^*,\xi)) = \mathcal{Q}_\beta(\widehat{J}(u^*,\xi)) + \Psi(u^*) - \Psi(u^*) \leq \mathcal{Q}_\beta(\widehat{J}(0,\xi)) - \Psi(u^*) \leq \mathcal{Q}_\beta(\widehat{J}(0,\xi)),$$

where the superquantile $\mathcal{Q}_\beta$ is defined in (3.4.2). Moreover, (3.3.5) and (3.2.28) yield $\widehat{J}(0,\xi) = (1/2)\|0 - y_d\|_H^2$, which is a constant. Since $\mathcal{Q}_\beta(z) = z$ for all constants $z \in \mathbb{R}$ (the superquantile is a coherent measure of risk [294, pp. 279 and 291]), we have $\mathcal{Q}_\beta(\widehat{J}(0,\xi)) = (1/2)\|y_d\|_H^2$. We have shown that $t^* \in T_{\mathrm{ad}} = [0, (1/2)\|y_d\|_H^2]$, which is also independent of $\beta \in (0,1)$.

3.4.1 Expected Value of the SAA Optimal Value

We derive bounds on the expected value of the SAA problem's optimal value (3.4.5).

Proposition 3.4.1. *Let $\beta \in (0,1)$ and let Assumptions 3.2.1* (a), *3.2.1* (c), *3.2.1* (d), *3.2.3* (a), *and 3.2.5* (a) *hold. Suppose that $\widehat{J}(u,\xi)$ is integrable for all u in a convex neighborhood of U_{ad}. Let (u^*,t^*) be an optimal solution of* (3.4.3) *with $t^* \in T_{ad}$, and for each $\omega \in \Omega^\star$, let $(u_N^*(\omega), t_N^*(\omega))$ be a minimizer of* (3.4.5). *Suppose that* $\mathrm{Prob}(\widehat{J}(u^*,\xi) = t^*) = 0$, *and that $R(U_{ad}) = \sup_{u\in U_{ad}} \|u - u^*\|_U$ and $R(T_{ad}) = \sup_{t\in T_{ad}} |t - t^*|$ are finite. Then $\mathbb{E}[f_N(u_N^*, t_N^*)] \leq f(u^*,t^*)$ and*

$$\begin{aligned} f(u^*,t^*) \leq \mathbb{E}[f_N(u_N^*,t_N^*)] &+ \tfrac{1}{\sqrt{N}}\big(\tfrac{\beta}{1-\beta}\big)^{1/2} R(T_{ad}) \\ &+ \tfrac{R(U_{ad})}{\sqrt{N}} \min\big\{ \tfrac{1}{1-\beta}\mathbb{E}[\|\nabla_u \widehat{J}(u^*,\xi)\|_U^2]^{1/2}, \tfrac{1}{(1-\beta)^{1/2}} \|\nabla_u \widehat{J}(u^*,\xi)\|_{L^\infty(\Xi;U)} \big\}, \end{aligned} \tag{3.4.6}$$

where f is defined in (3.4.3) *and f_N in* (3.4.5).

We establish Proposition 3.4.1 using Lemmas 3.4.2–3.4.6. We identify $(U\times\mathbb{R})^*$ with $U^*\times\mathbb{R}$.

Lemma 3.4.2. *Let $\xi \in \Xi$ be arbitrary, and let $\widehat{J}(\cdot,\xi) : U \to \mathbb{R}$ be convex, continuous, and Gâteaux differentiable at $(\bar{u},\bar{t}) \in U\times\mathbb{R}$. We define $\mathsf{F}(\cdot,\cdot;\xi) : U\times\mathbb{R} \to \mathbb{R}$ by $\mathsf{F}(u,t;\xi) = t + (1-\beta)^{-1}(\widehat{J}(u,\xi) - t))_+$. Then $\partial_{(u,t)}\mathsf{F}(\bar{u},\bar{t};\xi)$ is given by* (3.4.7).

[3] Indeed, for $t < 0$, $\beta \in (0,1)$ and $Z \in L^1(\Omega)$ with $Z \geq 0$, we have $t + (1/(1-\beta))\mathbb{E}[(Z-t)_+] = t + (1/(1-\beta))\mathbb{E}[Z-t] = t(1 - 1/(1-\beta)) + (1/(1-\beta))\mathbb{E}[Z] > (1/(1-\beta))\mathbb{E}[Z] = 0 + (1/(1-\beta))\mathbb{E}[(Z-0)_+] \geq \mathcal{Q}_\beta(Z)$.

Proof. We define $\mathsf{f}_1 : U \times \mathbb{R} \to \mathbb{R}$ by $\mathsf{f}_1(u,t) = t$, $\varphi : \mathbb{R} \to \mathbb{R}$ by $\varphi(z) = (1-\beta)^{-1}(z)_+$, and $\mathsf{f}_2 : U \times \mathbb{R} \to \mathbb{R}$ by $\mathsf{f}_2(u,t) = \widehat{J}(u,\xi) - t$. We have $\mathsf{F}(u,t;\xi) = \mathsf{f}_1(u,t) + \varphi(\mathsf{f}_2(u,t))$ for all $(u,t) \in U \times \mathbb{R}$. The function φ is convex and monotone, and f_1 and f_2 are convex and Gâteaux differentiable at $(\bar{u},\bar{t})$. Hence, $\partial_{(u,t)}\mathsf{f}_1(\bar{u},\bar{t}) = \{\,(0,1)\,\}$ and $\partial_{(u,t)}\mathsf{f}_2(\bar{u},\bar{t}) = \{\,(\mathrm{D}_u\widehat{J}(\bar{u},\xi),-1)\,\}$ [46, Prop. 2.125]. Combined with [161, Thm. 2 on p. 46], we get $\partial_{(u,t)}[\varphi(\mathsf{f}_2(\bar{u},\bar{t}))] = (\mathrm{D}_u\widehat{J}(\bar{u},\xi),-1)^*\partial\varphi(\mathsf{f}_2(\bar{u},\bar{t}))$. Now, the Moreau–Rockafellar theorem [46, Thm. 2.168] yields

$$\partial_{(u,t)}\mathsf{F}(\bar{u},\bar{t};\xi) = \partial_{(u,t)}\mathsf{f}_1(\bar{u},\bar{t}) + \partial_{(u,t)}[\varphi(\mathsf{f}_2(\bar{u},\bar{t}))] = \begin{bmatrix} \frac{1}{1-\beta}\partial(\widehat{J}(\bar{u},\xi)-\bar{t})_+\mathrm{D}_u\widehat{J}(\bar{u},\xi) \\ 1 - \frac{1}{1-\beta}\partial(\widehat{J}(\bar{u},\xi)-\bar{t})_+ \end{bmatrix}. \tag{3.4.7}$$

□

Lemma 3.4.3. *If the hypotheses of Proposition 3.4.1 hold, then the function F defined in* (3.4.4) *is Hadamard differentiable at (u^*,t^*).*

Proof. We define $\mathsf{f} : (U \times \mathbb{R}) \times \Xi \to \mathbb{R}$ by $\mathsf{f}(u,t,\xi) = t + (1-\beta)^{-1}(\widehat{J}(u,\xi) - t)_+$. Because $\widehat{J}$ is a convex Carathéodory function (see Assumption 3.2.1 (d)), f is one as well. Since $\widehat{J}(u,\xi)$ is integrable for all $u \in V$ for some convex neighborhood V of U_{ad}, $\mathsf{f}(u,t,\xi)$ is integrable for each $(u,t) \in V \times \mathbb{R}$. Moreover, $U \times \mathbb{R}$ is separable by Assumption 3.2.1 (a). Hence, F is continuous at (u^*,t^*) [161, Thm. 1 and Rem. 1 on p. 10] and $\partial F(u^*,t^*) = \mathbb{E}[\partial_{(u,t)}\mathsf{f}(u^*,t^*,\xi)]$ [161, Thm. 1 on p. 10].
Using the notation of Lemma 3.4.2, we have $\mathsf{f}(u,t,\xi) = \mathsf{F}(u,t;\xi)$ for all $(u,t,\xi) \in U \times \mathbb{R} \times \Xi$. We have $\partial(z)_+ = \{0\}$ if $z < 0$, $\partial(z)_+ = [0,1]$ if $z = 0$, and $\partial(z)_+ = \{1\}$ otherwise. Since $\mathrm{Prob}(\widehat{J}(u^*,\xi) = t^*) = 0$, (3.4.7) ensures that $\partial_{(u,t)}\mathsf{f}(u,t,\xi)$ is a singleton w.p. 1. Assumption 3.2.5 (a) and (3.2.13) yield $\mathbb{E}[\|\nabla_u\widehat{J}(u^*,\xi)\|_U] < \infty$. Combined with (3.4.7), we find that every measurable selection of $\partial_{(u,t)}\mathsf{f}(u,t,\xi)$ is Bochner integrable. Putting together the pieces, we find that $\partial F(u^*,t^*)$ is a singleton. Combined with the continuity of F at (u^*,t^*) and Assumption 3.2.1 (a), we conclude that F is Hadamard differentiable at (u^*,t^*) [46, Prop. 2.126]. □

Lemma 3.4.4. *If the hypotheses of Proposition 3.4.1 hold, then $\nabla_t F(u^*,t^*) = 0$ and*

$$(\nabla_u F(u^*,t^*), u_N^* - u^*)_U + \Psi(u) - \Psi(u^*) \geq 0 \quad \textit{for all} \quad u \in U_{ad}. \tag{3.4.8}$$

Proof. Lemma 3.4.3 ensures that the function F defined in (3.4.4) is Gâteaux differentiable at (u^*,t^*). Thus, $F(\cdot,t^*)$ and $F(u^*,\cdot)$ are Gâteaux differentiable at u^* and t^*, respectively. Assumption 3.2.1 (d) ensures the convexity of F, and U_{ad} is convex by Assumption 3.2.1 (a). Applying the arguments in the proof of [163, Thm. 3.1], and using Assumption 3.2.1 (c) and the fact that (u^*,t^*) is an optimal solution of (3.4.3), we obtain, for all $(u,t) \in U_{\mathrm{ad}} \times \mathbb{R}$,

$$(\nabla_u F(u^*,t^*), u - u^*)_U + \nabla_t F(u^*,t^*)(t-t^*) + \Psi(u) - \Psi(u^*) \geq 0.$$

Choosing $u = u^*$ yields $\nabla_t F(u^*,t^*)(t-t^*) \geq 0$ for all $t \in \mathbb{R}$. Consequently, $\nabla_t F(u^*,t^*) = 0$. Putting together the pieces, we obtain the assertions. □

Lemma 3.4.5 ([273, Thm. 10], [294, pp. 3, 92, and 276])**.** *We define $\mathsf{G} : L^1(\Omega) \times \mathbb{R} \to \mathbb{R}$ by $\mathsf{G}(Z,t) = t + (1-\beta)^{-1}\mathbb{E}[(Z-t)_+]$. For $\beta \in (0,1)$ and $Z \in L^1(\Omega)$, we have*

$$\begin{aligned} \arg\min_{t\in\mathbb{R}} \mathsf{G}(Z,t) &= \{\, t \in \mathbb{R} \colon\ \mathrm{Prob}(Z < t) \leq \beta \leq \mathrm{Prob}(Z \leq t) \,\} \\ &= \big[\, \inf_{\tau\in\mathbb{R}} \{\, \tau \colon\ \mathrm{Prob}(Z \leq \tau) \geq \beta \,\}, \sup_{\tau\in\mathbb{R}} \{\, \tau \colon\ \mathrm{Prob}(Z \leq \tau) \leq \beta \,\} \big]. \end{aligned} \tag{3.4.9}$$

Lemma 3.4.6. *Let the hypotheses of Proposition 3.4.1 hold. We define* $Z : \Omega \to \mathbb{R}$ *by* $Z(\omega) = \partial(\widehat{J}(u^*, \xi(\omega)) - t^*)_+$. *Then,* $1-(1-\beta)^{-1}Z$ *has zero mean and* $\mathbb{E}[(1-(1-\beta)^{-1}Z)^2] = \beta/(1-\beta)$. *Furthermore,*

$$\mathbb{E}[\|Z\nabla_u \widehat{J}(u^*,\xi) - \mathbb{E}[Z\nabla_u \widehat{J}(u^*,\xi)]\|_U^2] \leq \min\{(1-\beta)\|\nabla_u \widehat{J}(u^*,\xi)\|^2_{L^\infty(\Xi;U)}, \mathbb{E}[\|\nabla_u \widehat{J}(u^*,\xi)\|_U^2]\}.$$

Proof. Since $\mathrm{Prob}(\widehat{J}(u^*,\xi) = t^*) = 0$, the random variable Z is single-valued w.p. 1 and, hence, well-defined. Lemmas 3.2.11 and 3.4.4, and (3.4.7) imply that $0 = \nabla_t F(u^*,t^*) = \mathbb{E}[1-(1-\beta)^{-1}Z]$. Using Lemma 3.4.5, we obtain $\mathrm{Prob}(\widehat{J}(u^*,\xi) < t^*) \leq \beta \leq \mathrm{Prob}(\widehat{J}(u^*,\xi) \leq t^*)$. Combined with $\mathrm{Prob}(\widehat{J}(u^*,\xi) = t^*) = 0$, we find that $\mathrm{Prob}(Z = 0) = \mathrm{Prob}(\widehat{J}(u^*,\xi) \leq t^*) = \beta$. Consequently,

$$\mathbb{E}[(1 - \tfrac{1}{1-\beta}Z)^2] = \int_{Z=0} 1\mathrm{d}P(\omega) + \int_{Z=1} (1 - \tfrac{1}{1-\beta})^2 \mathrm{d}P(\omega) = \beta + (1-\beta)(1-\tfrac{1}{1-\beta})^2 = \tfrac{\beta}{1-\beta}.$$

Now, we derive the bounds on the expectation of $\|Z\nabla_u \widehat{J}(u^*,\xi) - \mathbb{E}[Z\nabla_u \widehat{J}(u^*,\xi)]\|_U^2$. Since U is a Hilbert space, we obtain

$$\mathbb{E}[\|Z\nabla_u \widehat{J}(u^*,\xi) - \mathbb{E}[Z\nabla_u \widehat{J}(u^*,\xi)]\|_U^2] \leq \mathbb{E}[\|Z\nabla_u \widehat{J}(u^*,\xi)\|_U^2].$$

Hölder's ensures $\mathbb{E}[\|Z\nabla_u \widehat{J}(u^*,\xi)\|_U^2] \leq \|\nabla_u \widehat{J}(u^*,\xi)\|^2_{L^\infty(\Xi;U)}\mathbb{E}[Z^2]$ and $Z \in \{0,1\}$ w.p. 1 yields $\mathbb{E}[\|Z\nabla_u \widehat{J}(u^*,\xi)\|_U^2] \leq \mathbb{E}[\|\nabla_u \widehat{J}(u^*,\xi)\|_U^2]$. Since $1-(1-\beta)^{-1}Z$ has zero mean and $Z \in \{0,1\}$, we obtain $\mathbb{E}[Z^2] = \mathbb{E}[Z] = 1-\beta$. Putting together the pieces, we obtain the bounds. □

Proof of Proposition 3.4.1. The estimate $\mathbb{E}[f_N(u_N^*, t_N^*)] \leq f(u^*,t^*)$ follows from [294, Prop. 5.6]. We adapt the proof of Proposition 3.2.18 to establish the upper bound. Lemma 3.6.1 ensures that $f_N(u_N^*, t_N^*, \cdot) : \Omega^\star \to \mathbb{R}$ (see (3.4.5)) is measurable. Lemma 3.4.3 implies that F defined in (3.4.4) is Hadamard differentiable at the optimal solution (u^*,t^*) of (3.4.3). Hence, $F(\cdot,t^*)$ and $F(u^*,\cdot)$ are Gâteaux differentiable at u^* and t^*, respectively.
Let us identify the dual of the Hilbert space $U \times \mathbb{R}$ with $U \times \mathbb{R}$. Let $(g^i, r^i) : \Omega^\star \to U \times \mathbb{R}$ be a measurable selection of $\partial_{(u,t)}[t + \frac{1}{1-\beta}(\widehat{J}(u,\xi^i) - t)_+]$ at (u^*,t^*) for $i = 1,\ldots,N$, which exists; see, e.g., [161, Thm. 3 on p. 11]. We define $(g_N, r_N) = (1/N)\sum_{i=1}^N (g^i, r^i)$. We have $(g_N, r_N) \in \partial_{(u,t)} F_N(u^*,t^*)$ w.p. 1 (see (3.4.4)). Using the definition of f and of f_N (see (3.4.5) and (3.4.3)), and that of F and of F_N (see (3.4.4)), the variational inequality (3.4.8), and the definition of (g_N, r_N), we obtain

$$\begin{aligned} f_N(u_N^*, t_N^*) &= F_N(u_N^*, t_N^*) + \Psi(u_N^*) \\ &\geq F_N(u^*,t^*) + \Psi(u_N^*) + (g_N, u_N^* - u^*)_U + r_N(t_N^* - t^*) \\ &\geq F_N(u^*,t^*) + \Psi(u^*) + (g_N - \nabla_u F(u^*,t^*), u_N^* - u^*)_U + r_N(t_N^* - t^*). \end{aligned}$$

Combined with the Cauchy–Schwarz inequality, $\|u_N^* - u^*\|_U \leq R(U_{\mathrm{ad}})$, $t_N^* \in T_{\mathrm{ad}}$ (see (3.4.5)), and $|t_N^* - t^*| \leq R(T_{\mathrm{ad}})$, we find that

$$\begin{aligned} f_N(u_N^*, t_N^*) \geq f(u^*,t^*) &+ F_N(u^*,t^*) - F(u^*,t^*) \\ &- \|\nabla_u F_N(u^*,t^*) - \nabla_u F(u^*,t^*)\|_U R(U_{\mathrm{ad}}) - |r_N| R(T_{\mathrm{ad}}). \end{aligned}$$

Rearranging and taking expectations yields

$$f(u^*,t^*) \leq \mathbb{E}[f_N(u_N^*, t_N^*)] + \mathbb{E}[\|g_N - \nabla_u F(u^*,t^*)\|_U] R(U_{\mathrm{ad}}) + \mathbb{E}[|r_N|] R(T_{\mathrm{ad}}). \tag{3.4.10}$$

Since $\mathbb{E}[\|\nabla_u \widehat{J}(u^*,\xi)\|_U] < \infty$ by Assumption 3.2.5 (a), Lemmas 3.2.11, 3.4.3 and 3.4.4, and (3.4.7) ensure $\mathbb{E}[g^i] = \nabla_u F(u^*,t^*)$ and $\mathbb{E}[r^i] = \nabla_t F(u^*,t^*) = 0$ for $i = 1,\ldots,N$. Combined with the fact that U is a Hilbert space, we deduce from (3.4.10) and the Cauchy–Schwarz inequality,

$$f(u^*,t^*) \leq +(1/\sqrt{N})\mathbb{E}[\|g^1 - \nabla_u F(u^*,t^*)\|_U^2]^{1/2} R(U_{\mathrm{ad}}) + (1/\sqrt{N})\mathbb{E}[|r^1|^2]^{1/2} R(T_{\mathrm{ad}}).$$

Together with Lemmas 3.4.2 and 3.4.6, and (3.4.7), we obtain the assertions. □

3.4.2 Discussion

We discuss the hypotheses, the implications, and the limitations of Proposition 3.4.1.
While $\|\nabla_u \widehat{J}(u^*,\xi)\|_{L^\infty(\Xi;U)} < \infty$ implies $\mathbb{E}[\|\nabla_u \widehat{J}(u^*,\xi)\|_U^2] < \infty$, $(1-\beta)^{1/2}\|\nabla_u \widehat{J}(u^*,\xi)\|_{L^\infty(\Xi;U)}$ in (3.4.6) may be smaller than $(1/(1-\beta))\mathbb{E}[\|\nabla_u \widehat{J}(u^*,\xi)\|_U^2]^{1/2}$ for $\beta \approx 1$. Since $\mathcal{Q}_\beta(Z) \to \|Z\|_{L^\infty(\Omega)}$ as $\beta \to 1$ for $Z \in L^\infty(\Omega)$ [294, Rem. 24], a dependence of $f(u^*,t^*) - \mathbb{E}[f_N(u_N^*,t_N^*)]$ on $1-\beta$ seems unavoidable. For $\beta = 0$, we have $\mathcal{Q}_\beta(Z) = \mathbb{E}[Z]$ for all $Z \in L^1(\Omega)$ [294, Rem. 24]. When choosing $\beta = 0$ in (3.4.6), and u^* is an unconstrained minimizer of the risk-neutral problem (3.2.1), we find that the estimate (3.2.25) formally recovers that in (3.4.6). However, the bound (3.4.6) is generally more conservative than (3.2.25) for $\beta = 0$.
The condition $\mathrm{Prob}(\widehat{J}(u^*,\xi) = t^*) = 0$ in Proposition 3.4.1 implies that the function F defined in (3.4.4) is Hadamard differentiable at (u^*,t^*) as we have shown in Lemma 3.4.3. Here, (u^*,t^*) is an optimal solution of (3.4.3). When $\mathrm{Prob}(\widehat{J}(u^*,\xi) = t^*) > 0$, then F is generally nonsmooth (see (3.4.7)). In this case, it might still be possible to establish similar bounds as those in (3.4.6), but our approach exploits the condition $\mathrm{Prob}(\widehat{J}(u^*,\xi) = t^*) = 0$. An alternative approach for studying risk-averse optimization problems using the superquantile is smoothing [191, sect. 4.1.1], [184, sect. 3.4.3]. Even though the risk-averse problem (3.4.3) is "Hadamard-smooth" under the hypotheses of Proposition 3.4.1, its SAA problem (3.4.5) is generally nonsmooth.
We relate the condition $\mathrm{Prob}(\widehat{J}(u^*,\xi) = t^*) = 0$ in Proposition 3.4.1 to the Hadamard differentiability of the β-superquantile $\mathcal{Q}_\beta$ at $\mathcal{J}(u^*)$, where $\mathcal{J}(u)(\omega) = \widehat{J}(u,\xi(\omega))$ and $\beta \in (0,1)$. The hypothesis $\mathrm{Prob}(\widehat{J}(u^*,\xi) = t^*) = 0$ implies that $\mathcal{Q}_\beta$ is Hadamard differentiable at $\mathcal{J}(u^*)$ [294, Ex. 6.19]. For $\beta \in (0,1)$ and $Z \in L^1(\Omega)$, the *β-quantile functional* of Z is defined by $\mathfrak{q}_\beta(Z) = \inf_{t\in\mathbb{R}} \{\, t\colon \mathrm{Prob}(Z \le t) \ge \beta \,\}$; see, e.g, [337, Ex. 4.2]. Let u^* be an optimal solution of (3.4.1). Then, (u^*,t^*) with $t^* = \mathfrak{q}_\beta(\mathcal{J}(u^*))$ is an optimal solution of (3.4.5) [273, Thm. 10], [294, p. 292]. If, furthermore, $\mathcal{Q}_\beta$ is Hadamard differentiable at $\mathcal{J}(u^*)$, then $\mathrm{Prob}(\widehat{J}(u^*,\xi) = t^*) = 0$ [294, Ex. 6.19]. To discuss the Hadamard differentiability of the composite risk function $\mathcal{Q}_\beta \circ \mathcal{J}$ at u^*, the continuity of $\mathcal{J}: U \to L^1(\Omega)$ may be required [279, Cor. 3.3].
To analyze the expected value of the SAA problem's optimal value of the superquantile minimization problem (3.4.1), we reformulated it as the stochastic problem (3.4.3). For this problem, the derivation of effective exponential tail bounds on the optimal control is more complicated or may be impossible because the cost function is not strongly convex as a mapping of (u,t), which are the optimization variables of (3.4.3); see also [184, pp. 62–63].

3.5 Conclusion and Discussion

We analyzed the SAA approach applied to stochastic convex optimal control problems posed in Hilbert spaces. For strongly convex problems, we derived exponential tail bounds for the optimal control of the SAA problem in section 3.2.1. For convex problems, we provided confidence intervals for the optimal value of the stochastic program in section 3.2.4. In section 3.2.5, we applied our findings to linear-quadratic control problems with convex regularization.
In section 3.3, we established reliable error bounds on the random optimal control of the discretized SAA problem, the SAA problem approximated by finite elements. For our analysis, we assumed the random diffusion coefficient κ is an element of $L^\infty(\Xi, C^1(\bar{\mathcal{D}}))$; see Assumption 3.3.1. Our result may be generalized assuming only $\kappa \in L^\infty(\Xi, C^t(\bar{\mathcal{D}}))$ for some $0 < t \le 1$. This condition has been used, for example, in [131, p. A2758] for stochastic approximation with adaptive mesh refinement.
The exponential tail bounds for the optimal controls relied on the assumption that exact minimizers of the SAA problem are computed. However, it is also possible to derive bounds for inexact minimizers using, for example, the notion of ε-optimal solutions or that used by Wachsmuth

and Rösch [333, eq. (2.1)] for deterministic control problems.
We analyzed the performance of the SAA approach for convex optimization problems. A starting point for the analysis of the scheme applied to nonlinear, nonconvex stochastic control problems posed in Hilbert spaces is the perturbation analysis by Bonnans and Shapiro [46, sect. 4.4.1].
For PDE-constrained optimization problems under uncertainty, an open task is the comparison of the SAA approach with, for example, stochastic approximation [246, 131, 128, 243], and schemes using stochastic collocation [189, 125, 314] and those utilizing low-rank tensors [123, 124, 29].
While we have not derived non-asymptotic sample size estimates for infinite-dimensional optimization problems which ensure that (approximate) optimal solutions of the SAA problem provide reliable ε-optimal solutions of the true counterpart, we conjecture that the approach by Shapiro [291] and Shapiro and Nemirovski [296] can be adapted for analyzing certain PDE-constrained optimization problems. The main idea is to construct a totally bounded subset of the feasible set U_{ad} that contains the minimizers of the stochastic problem and of its SAA.
We outline the construction of a totally bounded subset of U_{ad} for the stochastic linear-quadratic control problem discussed in section 3.3. Let us consider $\mathcal{D} = (-1,1)^d$, $\gamma = 0$ (see (3.3.5)), and

$$V_{\mathrm{ad}} = \{\, u \in H^1(\mathcal{D}) : u \in U_{\mathrm{ad}},\, \|u\|_{H^1(\mathcal{D})} \leq \tfrac{1}{\alpha} \sup_{u \in U_{\mathrm{ad}}} \|z(u,\xi)\|_{L^\infty(\Xi;H^1(\mathcal{D}))} + (|\mathfrak{l}| + \mathfrak{u})\|1\|_{L^2(\mathcal{D})} \,\},$$

where $\alpha > 0$, $U_{\mathrm{ad}} = \{\, u \in L^2(\mathcal{D}) : \mathfrak{l} \leq u \leq \mathfrak{u} \,\}$ with $-\infty < \mathfrak{l} < 0 < \mathfrak{u} < \infty$ (see (3.3.4)), and z is defined in (3.2.31). Since $\gamma = 0$, (3.3.5) implies $\Psi = 0$. Lemma 3.3.2 and Friedrichs' inequality ensure that V_{ad} is bounded and closed. The set V_{ad} also has finite diameter $R(V_{\mathrm{ad}}) \in (0,\infty)$ w.r.t. the $H^1(\mathcal{D})$-norm. Lemma 3.3.9 gives $u^* \in V_{\mathrm{ad}}$, where u^* is the optimal control of (3.2.27). Combined with $V_{\mathrm{ad}} \subset U_{\mathrm{ad}}$, we find that u^* is an optimal solution to $\min_{u \in V_{\mathrm{ad}}} F(u)$. Here, F is the expectation function defined in (3.2.3). Adapting the arguments in the proof of Lemma 3.3.9, we can show that each optimal solution of the SAA problem (3.2.2) corresponding to (3.2.27) is contained in V_{ad}. Hence, the minimizer of (3.2.2) defined by the data from section 3.3 is an optimal solution to the "reduced" SAA problem $\min_{u \in V_{\mathrm{ad}}} F_N(u)$. Here, F_N is the sample average function defined in (3.2.3).
Birman and Solomjak [40, Thm. 5.2], [41, Thm. 1.7] demonstrate that the ν-covering number of the closed $H^1(\mathcal{D})$-unit ball w.r.t. the $L^2(\mathcal{D})$-norm is proportional to $(1/\nu)^d$. Hence, the ν-covering number of V_{ad} w.r.t. the $L^2(\mathcal{D})$-norm is bounded by an absolute constant times $(R(V_{\mathrm{ad}})/\nu)^d$. To summarize, the set $V_{\mathrm{ad}} \subset U_{\mathrm{ad}}$ has an explicit bound on its covering numbers w.r.t. the $L^2(\mathcal{D})$-norm, and it contains the optimal solutions of the linear-quadratic problem (3.2.27) and of its SAA problem.
Two further "building-blocks" are required by the theory of Shapiro [291] and Shapiro and Nemirovski [296] (see also section 3.1): (a) For all u_1, $u_2 \in U_{\mathrm{ad}}$, the mean-zero random variable $\widehat{J}(u_2,\xi) - \widehat{J}(u_1,\xi) - \mathbb{E}[\widehat{J}(u_2,\xi) - \widehat{J}(u_1,\xi)]$ is sub-Gaussian with parameter $\sigma > 0$; and (b) $\widehat{J}(\cdot,\xi)$ is Lipschitz continuous with a constant $L(\xi) \geq 0$ that satisfies $\mathbb{E}[\exp(tL(\xi))] < \infty$ for all sufficiently small $t > 0$. Here, $\widehat{J}$ is the parameterized objective function defined in (3.2.28).
Concerning (a), Lemma 3.3.2 ensures that $\sup_{u \in U_{\mathrm{ad}}} \|\widehat{J}(u,\xi)\|_{L^\infty(\Xi;\mathbb{R})}$ is finite. Hence, $\widehat{J}(u_2,\xi) - \widehat{J}(u_1,\xi) - \mathbb{E}[\widehat{J}(u_2,\xi) - \widehat{J}(u_1,\xi)]$ is sub-Gaussian with parameter $4 \sup_{u \in U_{\mathrm{ad}}} \|\widehat{J}(u,\xi)\|_{L^\infty(\Xi;\mathbb{R})}$ for all u_1, $u_2 \in U_{\mathrm{ad}}$ [57, p. 9]. To establish (b), we apply Lemma 3.3.2 to obtain $L = \sup_{u \in U_{\mathrm{ad}}} \|\nabla_u \widehat{J}(u,\xi)\|_{L^\infty(\Xi;U)} < \infty$. Hence, the continuously differentiable function $\widehat{J}(\cdot,\xi)$ is Lipschitz continuous on U_{ad} with deterministic Lipschitz constant L for all $\xi \in \Xi$ [151, p. 9].
These observations may allow us to establish non-asymptotic sample size estimates for the "reduced" control problem $\min_{u \in V_{\mathrm{ad}}} F(u)$, and finally for the linear-quadratic optimization problem $\min_{u \in U_{\mathrm{ad}}} F(u)$ (see (3.2.27)). However, we leave the details for future work.

3.6 Supplementary Material

3.6.1 Measurability of Optimal Values and Optimal Solutions

We summarize statements on the measurability of optimal values of stochastic programs and of the set of optimal solutions, and prove Lemma 3.2.2. In this section, $(\Omega, \mathsf{F}, \mathsf{P})$ is a probability space.

Lemma 3.6.1. *Let V be a separable Banach space. If $\mathsf{f} : V \times \Omega \to \mathbb{R} \cup \{\infty\}$ is a random lower-semicontinuous, then $\mathsf{v}^* : \Omega \to \bar{\mathbb{R}}$ defined by $\mathsf{v}^*(\omega) = \inf_{x \in V} \mathsf{f}(x, \omega)$ is F-measurable, and $\mathsf{R}^* : \Omega \to V$ defined by $\mathsf{R}^*(\omega) = \{ x \in V : \mathsf{f}(x, \omega) \leq \mathsf{v}^*(\omega) \}$ is closed-valued. If, in addition, the image of R^* is nonempty, then R^* has a measurable selection.*

Proof. For each $\omega \in \Omega$, $\mathsf{f}(\cdot, \omega)$ is lower semicontinuous and, hence, $\mathsf{R}^*(\omega)$ is closed. The measurability of v^* follows from [66, Cor. VII-2] (see also [66, Lem. III.38 and p. 80], [213, Prop. 6.1], and [267, p. 225]), and that of R^* from [66, Thm. III.38 and p. 80]. The existence of a measurable selection is a result of the theorem on measurable selections; see, e.g., [66, Thms. III.22 and III.38] and [11, Thm. 8.1.3]. □

Lower-semicontinuous functions define random lower-semicontinuous functions [276, Ex. 14.30].

Lemma 3.6.2. *If V is a Banach space and $\mathsf{g} : V \to \mathbb{R} \cup \{\infty\}$ is lower-semicontinuous, then $\mathsf{f} : V \times \Omega \to \mathbb{R} \cup \{\infty\}$ defined by $\mathsf{f}(x, \omega) = \mathsf{g}(x)$ is random lower-semicontinuous.*

Proof. For each $t \in \mathbb{R}$, $\mathsf{g}^{-1}((-\infty, t])$ is closed and, hence, $\mathsf{g}^{-1}((-\infty, t]) \in \mathcal{B}(V)$. Consequently, $\{ (x, \omega) \in V \times \Omega : \mathsf{f}(x, \omega) \leq t \} = \{ x \in V : \mathsf{g}(x) \leq t \} \times \Omega \in \mathcal{B}(V) \otimes \mathsf{F}$. □

Sums of random lower-semicontinuous are random lower-semicontinuous [268, Prop. 2M], [276, Prop. 14.44], [98, Lem. 6.2], [170, p. 197].

Lemma 3.6.3. *If V is a Banach space and f_1, $\mathsf{f}_2 : V \times \Omega \to \mathbb{R} \cup \{\infty\}$ are random lower-semicontinuous, then $\mathsf{f}_1 + \mathsf{f}_2$ is random lower-semicontinuous.*

Proof. For each $\omega \in \Omega$, $\mathsf{f}_1(\cdot, \omega) + \mathsf{f}_2(\cdot, \omega)$ is lower-semicontinuous. The measurability of $\mathsf{f}_1 + \mathsf{f}_2$ follows from [284, Lem. 8.10] and the fact that $(V \times \Omega, \mathcal{B}(V) \otimes \mathcal{F})$ is a measurable space. □

Proof of Lemma 3.2.2. We apply Lemma 3.6.1 to $\min_{u \in U} F_N(u, \omega) + \Psi(u) + I_{U_{\mathrm{ad}}}(u)$, which has the same optimal value and the same set of optimal solutions as (3.2.2). Here, $I_{U_{\mathrm{ad}}} : U \to \mathbb{R} \cup \{\infty\}$ is the indicator function of U_{ad}. According to Assumption 3.2.1 (a) and [46, Ex. 2.67], $I_{U_{\mathrm{ad}}}$ is proper, convex and lower-semicontinuous. Since $\xi^i(\Omega^\star) \subset \Xi$ for $i = 1, \ldots, N$, Assumption 3.2.1 (d) implies that $F_N(\cdot, \omega)$ (see (3.2.3)) is continuous for each $\omega \in \Omega^\star$, and $F_N(u, \cdot) : \Omega^\star \to \mathbb{R}$ is measurable for each $u \in U$ [169, Lems. 1.7 and 1.12]. Hence, F_N is a Carathéodory function. Combined with Assumption 3.2.1 (c) and Lemmas 3.6.2 and 3.6.3, we find that $U \times \Omega^\star \ni (u, \omega) \mapsto F_N(u, \omega) + \Psi(u) + I_{U_{\mathrm{ad}}}(u)$ is random lower-semicontinuous. Thus, Lemma 3.6.1 implies the assertions. □

3.6.2 Exponential Tail Bounds for Hilbert Space-Valued Random Variables

We restate and prove Theorem 3.2.16. Throughout the section, $(\Omega, \mathsf{F}, \mathsf{P})$ is a probability space.

Theorem 3.2.16. *Let $(\Omega, \mathsf{F}, \mathsf{P})$ be a probability space and let H be a separable Hilbert space. Suppose that $\mathsf{Z}_i : \Omega \to \mathsf{H}$ for $i = 1, 2, \ldots$ are independent, mean-zero random variables such that $\mathbb{E}[\exp(\tau^{-2} \|\mathsf{Z}_i\|_{\mathsf{H}}^2)] \leq \mathrm{e}$ for some $\tau > 0$. Then, for each $N \in \mathbb{N}$ and every $\varepsilon \geq 0$,*

$$\mathrm{Prob}(\|\mathsf{S}_N / N\|_{\mathsf{H}} \geq \varepsilon) \leq 2 \exp(-\tau^{-2} \varepsilon^2 N / 3), \tag{3.6.1}$$

where $\mathsf{S}_N = \mathsf{Z}_1 + \cdots + \mathsf{Z}_N$. *If, in addition,* $\|\mathsf{Z}_i\|_{\mathsf{H}} \leq \tau$ *w.p.* 1 *for* $i = 1, 2, \ldots$, *then* $\mathrm{Prob}(\|\mathsf{S}_N/N\|_{\mathsf{H}} \geq \varepsilon) \leq 2\exp(-\tau^{-2}\varepsilon^2 N/2)$.

We apply the following statements to prove Theorem 3.2.16.

Theorem 3.6.4 ([259, Thm. 3], [356, Thm. 3.3.4])**.** *Let* $N \in \mathbb{N}$ *and let* H *be a separable Hilbert space. Suppose that* $\mathsf{Z}_i : \Omega \to \mathsf{H}$ *for* $j = 1, \ldots, N$ *are independent, mean-zero, and measurable. Then, for all* $\lambda \geq 0$, $\mathbb{E}[\cosh(\lambda\|\mathsf{Z}_1 + \cdots + \mathsf{Z}_N\|_{\mathsf{H}})] \leq \prod_{i=1}^{N} \mathbb{E}[\exp(\lambda\|\mathsf{Z}_i\|_{\mathsf{H}}) - \lambda\|\mathsf{Z}_i\|_{\mathsf{H}}]$.

Lemma 3.6.5. *If* $(\mathsf{V}, \|\cdot\|_{\mathsf{V}})$ *is a Banach space, and* $\mathsf{Z} : \Omega \to \mathsf{V}$ *is strongly measurable such that* $\mathbb{E}[\exp(\sigma^{-2}\|\mathsf{Z}\|_{\mathsf{V}}^2)] \leq \mathrm{e}$ *for some* $\sigma > 0$, *then*

$$\mathbb{E}[\exp(\lambda\|\mathsf{Z}\|_{\mathsf{V}}) - \lambda\|\mathsf{Z}\|_{\mathsf{V}}] \leq \exp(3\lambda^2\sigma^2/4) \quad \textit{for all} \quad \lambda \in \mathbb{R}_+. \tag{3.6.2}$$

Proof. The proof is inspired by that of [294, Prop. 7.72]. To establish (3.6.2), we distinguish whether $\lambda \in [0, 4/(3\sigma)]$ or $\lambda \in (4/(3\sigma), \infty)$.
We consider the case $\lambda \in [0, 4/(3\sigma)]$. We have $9\lambda^2\sigma^2/16 \leq 1$. Hence $\mathbb{R}_{\geq 0} \ni s \mapsto s^{9\lambda^2\sigma^2/16}$ is concave. For all $s \in \mathbb{R}$, we have $\exp(s) \leq s + \exp(9s^2/16)$; see, e.g., [294, p. 449]. Combined with Jensen's inequality and $\mathbb{E}[\exp(\|\mathsf{Z}\|_{\mathsf{V}}^2/\sigma^2)] \leq \mathrm{e}$, we obtain

$$\mathbb{E}[\mathrm{e}^{\lambda\|\mathsf{Z}\|_{\mathsf{V}}} - \lambda\|\mathsf{Z}\|_{\mathsf{V}}] \leq \mathbb{E}[\mathrm{e}^{9\lambda^2\|\mathsf{Z}\|_{\mathsf{V}}^2/16}] \leq \mathbb{E}[\mathrm{e}^{\|\mathsf{Z}\|_{\mathsf{V}}^2/\sigma^2}]^{9\lambda^2\sigma^2/16} \leq \mathrm{e}^{9\lambda^2\sigma^2/16} \leq \mathrm{e}^{3\lambda^2\sigma^2/4}. \tag{3.6.3}$$

Now, we consider the case $\lambda \geq 4/(3\sigma)$. We have $2/3 \leq \sigma\lambda/2$. Consequently, $4/9 \leq \sigma^2\lambda^2/4$ and $2/3 \leq 3\sigma^2\lambda^2/8$. For all $s \in \mathbb{R}$, Young's inequality yields

$$\lambda s = [(3/4)^{1/2}\lambda][(4/3)^{1/2}s] \leq 3\lambda^2\sigma^2/8 + 2s^2/(3\sigma^2).$$

Combined with Jensen's inequality, $\mathbb{E}[\exp(\|\mathsf{Z}\|_{\mathsf{V}}^2/\sigma^2)] \leq \mathrm{e}$, and $2/3 \leq 3\sigma^2\lambda^2/8$, we get

$$\mathbb{E}[\mathrm{e}^{\lambda\|\mathsf{Z}\|_{\mathsf{V}}} - \lambda\|\mathsf{Z}\|_{\mathsf{V}}] \leq \mathbb{E}[\mathrm{e}^{\lambda\|\mathsf{Z}\|_{\mathsf{V}}}] \leq \mathrm{e}^{3\lambda^2\sigma^2/8}\mathbb{E}[\mathrm{e}^{2\|\mathsf{Z}\|_{\mathsf{V}}^2/(3\sigma^2)}] \leq \mathrm{e}^{3\lambda^2\sigma^2/8+2/3} \leq \mathrm{e}^{3\lambda^2\sigma^2/4}.$$

Together with (3.6.3), we obtain (3.6.2). □

Lemma 3.6.6. *If* $a, b > 0$, *then* $\min_{\lambda>0} -a\lambda + b\lambda^2 = -a^2/(4b)$.

Proof. The minimizer is $\lambda_* = a/(2b)$. Hence, $-a\lambda_* + b\lambda_*^2 = -a^2/(2b) + a^2/(4b) = -a^2/(4b)$. □

Proof of Theorem 3.2.16. We use a Chernoff-type approach to establish (3.6.1); see Chernoff [76, p. 496]. The second claim follows from an application of [256, Thm. 3.5]. Fix $\lambda > 0$, ε, $r \geq 0$, and $N \in \mathbb{N}$. Using Lemma 3.6.5 and $\mathbb{E}[\exp(\tau^{-2}\|\mathsf{Z}_i\|_{\mathsf{H}}^2)] \leq \mathrm{e}$, we find that

$$\prod_{i=1}^{N} \mathbb{E}[\exp(\lambda\|\mathsf{Z}_i\|_{\mathsf{H}}) - \lambda\|\mathsf{Z}_i\|_{\mathsf{H}}] \leq \prod_{i=1}^{N} \exp(3\lambda^2\tau^2/4) = \exp(3\lambda^2\tau^2 N/4).$$

Combined with Markov's inequality, Theorem 3.6.4, and $\exp(s) \leq 2\cosh(s)$ valid for all $s \in \mathbb{R}$, we obtain

$$\begin{aligned}\mathrm{Prob}(\|\mathsf{S}_N\|_{\mathsf{H}} \geq r) &\leq \exp(-\lambda r)\mathbb{E}[\exp(\lambda\|\mathsf{S}_N\|_{\mathsf{H}})] \leq 2\exp(-\lambda r)\mathbb{E}[\cosh(\lambda\|\mathsf{S}_N\|_{\mathsf{H}})] \\ &\leq 2\exp(-\lambda r + 3\lambda^2\tau^2 N/4).\end{aligned}$$

Minimizing the right-hand side over $\lambda > 0$ yields $\mathrm{Prob}(\|\mathsf{S}_N\|_{\mathsf{H}} \geq r) \leq 2\exp(-\tau^{-2}r^2/(3N))$ (see Lemma 3.6.6). Choosing $r = \varepsilon N$ implies (3.6.1).
If, in addition, $\|\mathsf{Z}_i\|_{\mathsf{H}} \leq \tau$, then [256, Thm. 3.5] yields $\mathrm{Prob}(\|\mathsf{S}_N/N\|_{\mathsf{H}} \geq \varepsilon) \leq 2\exp(-\tau^{-2}\varepsilon^2 N/2)$.
□

4 Exponential Tail Bounds for Multilevel Monte Carlo Mean Estimators in a Class of Smooth Banach Spaces

The Multilevel Monte Carlo (MLMC) mean estimator utilizes low and high fidelity models to estimate the expectation of a Banach space-valued random variable. We derive non-asymptotic, exponential bounds on the tail probabilities of the MLMC mean estimator for multilevel corrections that have sub-Gaussian behavior and take values in certain Banach spaces. The tail probability is the probability that the distance between the MLMC mean estimator and the true mean exceeds a prescribed accuracy. The tail bounds imply that the number of samples required to reliably estimate the mean via the MLMC estimator depend only moderately on the user-specified reliability. We develop our analysis for a class of smooth Banach spaces which includes, for example, all Hilbert spaces and all Sobolev space consisting of at least square-integrable functions. The approach also allows the mean of essentially bounded and continuous functions to be reliably estimated, even though the corresponding function spaces are nonsmooth. We demonstrate that our results apply to a class of linear elliptic partial differential equations with random inputs.

4.1 Introduction

Multilevel Monte Carlo (MLMC) methods can be used to estimate the mean of a Banach space-valued random variable, such as the mean of the solutions to partial differential equations (PDEs) with random inputs. MLMC methods utilize several low and high fidelity models, and are designed to perform many simulations with the low fidelity models, but relatively few with accurate approximations. The overall goal of the MLMC scheme is to save computational costs compared to the standard Monte Carlo (MC) estimator, while achieving the same performance guarantees. Typically, the accuracy of the MLMC mean estimator is measured using the mean-squared error [37, 134, 146].

We augment the existing error analysis of the MLMC mean estimator in that we derive non-asymptotic, exponential bounds on the tail probabilities of the MLMC mean estimator for certain Banach space-valued random variables. (An example of an exponential tail bound for the MLMC mean estimator is provided in (4.1.4).) The exponential bounds imply that the number of samples required to obtain a reliable mean estimate depend only moderately on the user-specified reliability. Our analysis is mainly inspired by and built on that developed by Juditsky and Nemirovski [166]. We establish the exponential tail bounds for the MLMC mean estimator by applying the exponential moment inequalities and the tail bounds established by Pinelis [256, 257, 258] and Pinelis and Sakhanenko [259]. For certain Sobolev space-valued random variables, we derive refined moment inequalities and exponential tail bounds.

We introduce the MLMC mean estimator and describe our contributions in detail. We choose $L \in \mathbb{N}$, a Banach space $(V, \|\cdot\|_V)$ and nested subspaces (V_ℓ) of V, and a probability space $(\Omega, \mathcal{F}, P)$. Moreover, let $X : \Omega \to V$ and let $X_\ell : \Omega \to V_\ell$ be Bochner integrable for $\ell = 1, \ldots, L$.

The MLMC mean estimator $\mathrm{E}^{\mathrm{ML}}[X_L]$ of $\mathbb{E}[X]$ is defined by

$$\mathrm{E}^{\mathrm{ML}}[X_L] = \sum_{\ell=1}^{L} \mathrm{E}^{N_\ell}[Y_\ell], \quad Y_\ell = X_\ell - X_{\ell-1}, \quad X_0 = 0, \tag{4.1.1}$$

where $\mathrm{E}^{N_\ell}[Y_\ell] = (1/N_\ell)\sum_{i=1}^{N_\ell} Y_{\ell,i}$ is the MC mean estimator. For each fixed ℓ, $Y_{\ell,i}$ has the same probability distribution as Y_ℓ for $i = 1, 2, \ldots$, and $Y_{\ell,i}$ for ℓ, $i = 1, 2, \ldots$ are independent [37, sect. 3], [134, sect. 1.3]. The functions Y_ℓ are called *multilevel corrections* [134, p. 8]. Here, the larger the index ℓ, the higher the fidelity of the model X_ℓ. We view $Y_{\ell,i}$ as V-valued random variables defined on a common probability space; see [44, pp. 148–149] for the standard construction of such a space.

For fixed $\varepsilon > 0$ and $0 < \delta \ll 1$, we are interested in determining the number of samples N_ℓ ($\ell = 1, \ldots, L$) such that

$$\mathrm{Prob}(\|\mathrm{E}^{\mathrm{ML}}[X_L] - \mathbb{E}[X]\|_V \geq \varepsilon) \leq \delta. \tag{4.1.2}$$

We refer to $\varepsilon \geq 0$ as *accuracy* and to $1-\delta \in (0,1)$ as *reliability*. When the multilevel corrections Y_ℓ ($\ell = 1, \ldots, L$) have sub-Gaussian tail behavior, we show that N_ℓ, for $\ell = 1, \ldots, L$, depends only logarithmically on $1/\delta$. In this case, values of δ, say $\delta = 10^{-8}$ or $\delta = 10^{-12}$, result in moderate values of N_ℓ. The tail bound (4.1.2) expresses the fact that, with a probability of at least $1 - \delta$, the realizations of the MLMC estimator are in an ε-ball about the true mean.

Before discussing our contributions in detail, we illustrate how to derive non-asymptotic, exponential bounds on $\mathrm{Prob}(\|\mathrm{E}^{\mathrm{ML}}[X_L] - \mathbb{E}[X]\|_V \geq \varepsilon)$ for $V = \mathbb{R}$ and fixed $\varepsilon > 0$. We assume that $|\mathbb{E}[X_L] - \mathbb{E}[X]| \leq \varepsilon/2$ for some $L \in \mathbb{N}$, and that the centered multilevel correction $Y_\ell - \mathbb{E}[Y_\ell]$ is sub-Gaussian with parameter $\tau_\ell > 0$ for $\ell = 1, \ldots, L$. Here, a random variable $\xi : \Omega \to \mathbb{R}$ is *sub-Gaussian with parameter* τ if $\tau \geq 0$ and $\mathbb{E}[\exp(\lambda\xi)] \leq \exp(\lambda^2\tau^2/2)$ for all $\lambda \in \mathbb{R}$ [57, p. 2]. In this case, [57, Lems. 1.3 and 1.7 (sect. 1.1)] and the definition of the MLMC mean estimator yield, for all $r \geq 0$,

$$\mathrm{Prob}(|\mathrm{E}^{\mathrm{ML}}[X_L] - \mathbb{E}[X_L]| \geq r) \leq 2\exp\left(\frac{r^2}{2\sum_{\ell=1}^{L}\tau_\ell^2/N_\ell}\right). \tag{4.1.3}$$

Combined with the triangle inequality, and the monotonicity of $\mathrm{Prob}(\cdot)$, we find that

$$\mathrm{Prob}(|\mathrm{E}^{\mathrm{ML}}[X_L] - \mathbb{E}[X]| \geq r + |\mathbb{E}[X_L] - \mathbb{E}[X]|) \leq 2\exp\left(\frac{r^2}{2\sum_{\ell=1}^{L}\tau_\ell^2/N_\ell}\right) \quad \text{for all} \quad r > 0.$$

When choosing $r = \varepsilon/2$ and using $|\mathbb{E}[X_L] - \mathbb{E}[X]| \leq \varepsilon/2$, we obtain

$$\mathrm{Prob}(|\mathrm{E}^{\mathrm{ML}}[X_L] - \mathbb{E}[X]| \geq \varepsilon) \leq 2\exp\left(\frac{\varepsilon^2}{8\sum_{\ell=1}^{L}\tau_\ell^2/N_\ell}\right). \tag{4.1.4}$$

In order for (4.1.2) to hold, we bound the right-hand side in (4.1.4) by δ, resulting in the requirement that N_ℓ ($\ell = 1, \ldots, L$) must satisfy

$$\sum_{\ell=1}^{L} \frac{\tau_\ell^2}{N_\ell} \leq \frac{\varepsilon^2}{8\ln(2/\delta)}. \tag{4.1.5}$$

We compare (4.1.3) with a tail bound obtained via the direct application of Tschebyshev's inequality. Tschebyshev's inequality yields, for all $r > 0$,

$$\mathrm{Prob}(|\mathrm{E}^{\mathrm{ML}}[X_L] - \mathbb{E}[X_L]| \geq r) \leq r^{-2}\sum_{\ell=1}^{L}\sigma_\ell^2/N_\ell,$$

where $\sigma_\ell^2 = \mathbb{E}[(Y_\ell - \mathbb{E}[Y_\ell])^2]$. In order to obtain (4.1.2) via this tail bound, we would need to choose N_ℓ ($\ell = 1, \dots, L$) according to

$$\sum_{\ell=1}^{L} \frac{\sigma_\ell^2}{N_\ell} \leq \frac{\varepsilon^2 \delta}{4}. \tag{4.1.6}$$

Both σ_ℓ and τ_ℓ are problem-dependent constants, and $\sigma_\ell^2 \leq \tau_\ell^2$ [57, Lem. 1.2, p. 3]. Moreover, if the centered multilevel corrections are Gaussian, we have $\sigma_\ell^2 = \tau_\ell^2$ [57, p. 2]. The estimate (4.1.5) depends on $1/\ln(2/\delta)$, whereas the bound (4.1.6) depends linearly on δ. Consequently, the former bound yields a less restrictive condition on N_ℓ ($\ell = 1, \dots, L$) than the latter one does for small $\delta \in (0, 1)$ and $\sigma_\ell^2 \approx \tau_\ell^2$. Furthermore, the exponential rate of the tail bound (4.1.3) is optimal under the stated assumptions [57, p. 19].

The MLMC method becomes meaningful if the computational cost $C_\ell > 0$ for sampling Y_ℓ and the sub-Gaussian parameters τ_ℓ decrease as ℓ increases, and bounds on the bias term $|\mathbb{E}[X_L] - \mathbb{E}[X]|$ are available. We use a simple model that allows for a complexity analysis of the MLMC scheme. The model by Giles [133, p. 609] is adapted in section 4.3. The simple model is more restrictive than that in [133, p. 609], but allows for explicit computations. We assume the existence of α, β, $\gamma > 0$ such that for $\ell \geq 2$, we have $C_\ell > 0$ and

$$|\mathbb{E}[X_\ell] - \mathbb{E}[X]| \leq (1/2)^\alpha |\mathbb{E}[X_{\ell-1}] - \mathbb{E}[X]|, \quad \tau_\ell^2 \leq (1/2)^\beta \tau_{\ell-1}^2, \quad \text{and} \quad C_{\ell-1} \leq (1/2)^\gamma C_\ell. \tag{4.1.7}$$

If $L \in \mathbb{N}$ and $L \geq 1 + (1/\alpha)\log_2((2/\varepsilon)|\mathbb{E}[X_1] - \mathbb{E}[X]|)$, then (4.1.7) ensures $|\mathbb{E}[X_L] - \mathbb{E}[X]| \leq \varepsilon/2$. The remaining goal is to choose N_ℓ ($\ell = 1, \dots, N$) such that the cost of the MLMC mean estimator $\sum_{\ell=1}^{L} N_\ell C_\ell$ is minimized subject to the tail bound (4.1.2). Since (4.1.5) when combined with $|\mathbb{E}[X_L] - \mathbb{E}[X]| \leq \varepsilon/2$ ensures (4.1.2), we can obtain N_ℓ ($\ell = 1, \dots, N$) as an (approximate) solution to

$$\min_{\substack{N_\ell \in \mathbb{N} \\ \ell = 1, \dots, L}} \sum_{\ell=1}^{L} N_\ell C_\ell \quad \text{s.t.} \quad \sum_{\ell=1}^{L} \frac{\tau_\ell^2}{N_\ell} \leq \frac{\varepsilon^2}{8\ln(2/\delta)}. \tag{4.1.8}$$

When approximating the constraints $N_\ell \in \mathbb{N}$ with $N_\ell \in (0, \infty)$, the optimal solution of the relaxation is

$$N_\ell = c(\tau_\ell^2/C_\ell)^{1/2}(8\ln(2/\delta)/\varepsilon^2) \quad \text{for} \quad \ell = 1, \dots, L \quad \text{with} \quad c = \sum_{\ell=1}^{L} (\tau_\ell^2 C_\ell)^{1/2};$$

cf. [134, p. 262]. Combined with the conditions in (4.1.7), we obtain $N_\ell \leq (1/2)^{(\beta+\gamma)/2} N_{\ell-1}$ for $\ell = 2, \dots, L$. Hence, the sample sizes depend only logarithmically on $1/\delta$, whereas the bound (4.1.6) would yield a linear dependence on $1/\delta$. Moreover, the sample size estimates decrease q-linearly with rate $(1/2)^{(\beta+\gamma)/2}$. The corresponding cost of the MLMC mean estimator is $(8\ln(2/\delta)/\varepsilon^2)(\sum_{\ell=1}^{L}(\tau_\ell^2 C_\ell)^{1/2})^2$; in contrast the cost of the MLMC mean estimator is $(4/(\delta\varepsilon^2))(\sum_{\ell=1}^{L}(\sigma_\ell^2 C_\ell)^{1/2})^2$ when using the bound (4.1.6) instead of (4.1.5), and assuming $\sigma_\ell^2 \leq (1/2)^\beta \sigma_{\ell-1}^2$ in (4.1.7) instead of $\tau_\ell^2 \leq (1/2)^\beta \tau_{\ell-1}^2$.

Our primary goals are to extend the validity of the exponential tail bound (4.1.3) to spaces V other than $V = \mathbb{R}$, and to analyze the computational complexity of the resulting MLMC scheme. We require that either the space V or the spaces V_ℓ ($\ell = 1, \dots, L$) are 2-uniformly smooth, or that they are 2-uniformly smooth after an equivalent renorming. For example, all Hilbert spaces and each Sobolev spaces consisting of at least square integrable functions are 2-uniformly smooth. Our results also apply to mean estimation when V is "nonsmooth." For instance, if V is the space of either the essentially bounded or continuous functions, the corresponding

function spaces are nonreflexive, and hence they cannot be equipped with a 2-uniformly smooth norm [233, Cor. 1.1]. In this case, we exploit the fact that the spaces V_ℓ ($\ell = 1, \dots, L$) are often finite-dimensional spaces in the context of uncertainty quantification with PDEs. We show that certain finite element spaces can be equipped with an equivalent 2-uniformly smooth norm and, moreover, we establish that the 2-uniform smoothness constant depends only logarithmically on the dimension of these finite-dimensional spaces.
We express sub-Gaussianity of the multilevel correction Y_ℓ through the condition

$$\mathbb{E}[\exp(\|Y_\ell - \mathbb{E}[Y_\ell]\|_V^2/\tau_\ell^2)] \leq \mathrm{e} \quad \text{for some} \quad \tau_\ell > 0. \tag{4.1.9}$$

The condition (4.1.9) and its variants are primarily used in the literature on stochastic programming [99, p. 679], [243, eq. (2.50)], [138, pp. 1035–1036], [294, eq. (5.347)]. The assumption (4.1.9) is fulfilled for V-valued random variables that are, for example, real-valued sub-Gaussian [57, Lem. 1.9, p. 9], Gaussian (due to the Landau–Shepp–Fernique theorem), certain Besov "priors" [87, Thm. 5], γ-sub-Gaussian [118, Thm. 3.4], sub-Gaussian random series [56, Thm. 1.10.3], or essentially bounded. We show that multilevel corrections corresponding to the solutions of linear elliptic PDEs with uniformly bounded random diffusion coefficient fulfill (4.1.9), such as those considered in [16, sect. 3]. We also demonstrate that solutions to elliptic PDEs with log-normal random diffusion coefficients may violate (4.1.9).

Related Work

The accuracy of MLMC mean estimators is typically quantified using the mean-squared error [134, 16, 37, 38, 182, 183]. A small mean-squared error yields upper bounds on tail probabilities via Tschebyshev's inequality. These bounds are polynomials as functions of the variance of the estimator's variance and the inverse of the accuracy. However, in order to obtain (4.1.2) with high confidence $1 - \delta$, these bounds would typically require a large number of samples, rendering the task of reliably estimating the mean intractable. MLMC estimators using quasi-Monte Carlo techniques to approximate the expectations of the multilevel corrections are presented in [203, 204].
The complexity of achieving certain mean-squared errors for the MLMC mean estimators is analyzed, for example, in [134, 16, 37, 182, 311]. The authors of [82] and of [143] provide asymptotic confidence intervals of MLMC mean estimators for real-valued quantities of interests, using the central limit theorem. Approximate confidence intervals of MLMC mean estimators for real-valued random variables are also developed [104]. However, asymptotic and approximate confidence intervals, constructed with the central limit theorem, tend to be optimistic and unreliable when the sample size is small. Our confidence regions are valid for real-valued random variables as well, are non-asymptotic, and are optimal up to problem-independent, moderate constants.
Kebaier and Lelong [176] have developed a MLMC method using importance sampling and prove a strong law of large numbers and a central limit theorem for this MLMC method. Jourdain and Kebaier [165] have established non-asymptotic confidence bounds for the MLMC Euler method for Lipschitz continuous real-valued functions of solutions to a certain class of stochastic differential equations. Our approach for deriving exponential tail bounds is similar. However, our results are not restricted to the MLMC Euler estimator, and are valid for random variables other than real-valued ones.
Most contributions, such as [16, 182, 311], focus on MLMC methods for real-valued and Hilbert space-valued random variables. Heinrich [146, sect. 4] provides an error and cost analysis for the mean-squared error of MLMC methods in separable Banach spaces of (Rademacher-)type p for $1 \leq p \leq 2$. Banach spaces that are 2-uniformly smooth are of type 2 [212, Lem. 2.2]. Examples

of type 2 spaces that are not 2-uniformly smooth are provided in [262, Chap. 12]. Daun and Heinrich [88, 89] study the complexity of multilevel algorithms for type p Banach space-valued random variables using classical error measures.
To derive non-asymptotic exponential tail bounds for the MLMC mean estimators, we apply the exponential moment inequalities and the tail bounds established in [256, 259, 258]. Exponential tail bounds are derived in [255, Thm. 1] and in [256] for bounded martingales that take values in 2-uniformly smooth spaces, and in [259, Thm. 3] and [356, Thm. 3.3.4] for martingales and independent random variables with values in Hilbert spaces, respectively. Exponential moment inequalities and tail bounds for general Banach spaces can be found in [261, Lem. 2.7], [309, Thms. 3 and 4]. However, these inequalities depend on an unspecified constant.
Confidence regions for (mean) estimators can also be established via Berry–Esséen-type inequalities, an approach which is used in [221, Chap. 4] to derive confidence intervals for sample means of real-valued random variables. Berry–Esséen inequalities provide non-asymptotic rates for the convergence asserted by the central limit theorem, under suitable assumptions. Lord, Powell, and Shardlow [221] also outline how the Berry–Esséen inequality for Hilbert space-valued random variables can be used to derive confidence regions for MC mean estimators [221, p. 424]. Ben Alaya and Kebaier [19] have established a central limit theorem and a Berry–Esséen-type bound for the MLMC Euler method. For Hilbert space-valued random variables, Berry–Esséen inequalities are derived, for example, in [355, 201]. However, these inequalities depend on an unspecified constant.

Outline

In section 4.2, we introduce some notation and Orlicz spaces, which allow for a metric characterization of the condition (4.1.9). We define two notions of smoothness of Banach spaces—2-uniform smoothness of 2-quasi-smoothness—in section 4.2.2, and provide several examples of such spaces. Bounds on the second moment of martingale-differences and sums of random averages are presented in section 4.2.3. We conclude section 4.2 with stating exponential tail bounds. The proofs of the statements from sections 4.2.3 and 4.2.4 are presented in section 4.7. In section 4.3, we use the above results to derive exponential tail bounds for the MLMC mean estimator, and demonstrate that the MLMC mean estimator is computationally cheaper than the sample mean, under suitable assumptions. Whereas the space of essentially bounded functions and that of continuous functions fail to be 2-uniformly smooth, we show that certain finite element spaces are 2-quasi-smooth subspaces in section 4.4. Our theory is applied in section 4.5 to linear elliptic PDEs with random inputs. We summarize our contributions and outline possible further research directions in section 4.6.

4.2 Notation and Preliminaries

Following [37, p. 587], we define $\mathrm{Cost}(\cdot) : L^0(\Omega; V) \to [0, \infty)$ as an abstract, measurable evaluation cost. For $W \in L^0(\Omega; V)$, $\mathrm{Cost}(W)$ is measurable [159, Cor. 1.1.24]. For two random variables a, $b : \Omega \to \mathbb{R}_+$, $a(\omega) \lesssim b(\omega)$ means $a(\omega) \leq Cb(\omega)$ for some constant $C > 0$ that is independent of $\omega \in \Omega$ and $b(\omega)$, and $a(\omega) \simeq b(\omega)$ abbreviates $a(\omega) \lesssim b(\omega)$ and $-a(\omega) \lesssim -b(\omega)$, as in [70, p. 324]. For $x \in \mathbb{R}$, $\lceil x \rceil_{\mathbb{N}} \in \mathbb{N}$ is the smallest number such that $x \leq \lceil x \rceil_{\mathbb{N}}$. For each martingale or martingale-difference $(Z_i)_{i \in \mathbb{N}_0} \subset L^1(\Omega; V)$ that is adapted to some filtration $(\mathcal{F}_i)_{i \in \mathbb{N}_0} \subset \mathcal{F}$, we set $Z_0 = 0$ and $\mathcal{F}_0 = \{\emptyset, \Omega\}$. Throughout, we use the following facts: (i) If $(V, \|\cdot\|_V)$ is a (reflexive) separable Banach space and $|||\cdot|||_V$ is an equivalent norm on V, then $(V, |||\cdot|||_V)$ is a (reflexive) separable Banach space. (ii) If $X \in L^0(\Omega; V)$, then $\|X\|_V$ is a real-valued random variable [159, Cor. 1.1.24]. (iii) Partial sums of independent, mean-zero,

Bochner integrable, V-valued random variables define a (stopped) martingale adapted to the natural filtration [159, Ex. 3.1.4].

4.2.1 Orlicz Spaces

We define (Fenchel–)Orlicz spaces and equip them with the Luxemburg norm. The Orlicz spaces are used to characterize certain "light-tailed" V-valued random variables, and we make use of the triangle inequality to analyze MLMC mean estimators.
We define the Young function $\psi : \mathbb{R} \to \mathbb{R}_+$ on $\mathbb{R}$ by $\psi(x) = (\mathrm{e}^{x^2} - 1)/(\mathrm{e} - 1)$ [320, Def. 1.1]. The Orlicz space $L_\psi(\Omega; V) = L_{\psi(\|\cdot\|_V)}(\Omega, \mathcal{F}, P; V)$ is the set of functions $Z \in L^0(\Omega; V)$ such that there exists $\tau > 0$ with $\mathbb{E}[\psi(\|Z\|_V/\tau)] < \infty$ [320, Def. 1.2]. Here, $L^0(\Omega; V)$ is the set of strongly measurable functions from Ω to V (see p. viii).
We define the Luxemburg norm $\|\cdot\|_{L_\psi(\Omega;V)}$ on $L_\psi(\Omega; V)$ by

$$\|Z\|_{L_\psi(\Omega;V)} = \inf_{\tau>0} \left\{ \tau : \mathbb{E}[\psi(\|Z\|_V/\tau)] \le 1 \right\} = \inf_{\tau>0} \left\{ \tau : \mathbb{E}[\exp(\|Z\|_V^2/\tau^2)] \le \mathrm{e} \right\}. \tag{4.2.1}$$

The Orlicz space $(L_\psi(\Omega; V), \|\cdot\|_{L_\psi(\Omega;V)})$ is a Banach space [320, Cor. 2.23].
If $Z \in L_\psi(\Omega; V) \setminus \{0\}$, then the infimum in (4.2.1) is attained [320, Lem. 2.17]. The definition of the Luxemburg norm ensures $\|Z\|_{L_\psi(\Omega;V)} \le \|Z\|_{L^\infty(\Omega;V)}$ if $Z \in L^\infty(\Omega; V)$ and, moreover, $\|Z_1\|_{L_\psi(\Omega;V)} \le \|Z_2\|_{L_\psi(\Omega;V)}$ if, w.p. 1, $\|Z_1\|_V \le \|Z_2\|_V$ and Z_1, $Z_2 \in L_\psi(\Omega; V)$.
The term "light-tailed" is motivated by the following fact: if $Z \in L^0(\Omega; V)$, then $Z \in L_\psi(\Omega; V)$ if and only if $\mathrm{Prob}(\|Z\|_V \ge \varepsilon) \le c_1 \exp(-\varepsilon^2/c_2^2)$ for all $\varepsilon > 0$ and some c_1, $c_2 > 0$; see, e.g., [57, pp. 55–56].

Lemma 4.2.1. *If $Z \in L_\psi(\Omega; V)$ and $\widetilde{Z} \in V$, then $\mathbb{E}[\|Z - \widetilde{Z}\|_V^2] \le \|Z - \widetilde{Z}\|_{L_\psi(\Omega;V)}^2$ and $\|Z - \mathbb{E}[Z]\|_{L_\psi(\Omega;V)} \le 2\|Z - \widetilde{Z}\|_{L_\psi(\Omega;V)}$.*

Proof. The first estimate follows from an application of Jensen's inequality; see also [243, p. 1584]. If $Z = \widetilde{Z}$, then $\mathbb{E}[Z] = \widetilde{Z} = Z$. Now, let $Z \ne \widetilde{Z}$. The triangle inequality and Jensen's inequality yield $\|Z - \mathbb{E}[Z]\|_V^2 \le 2\|Z - \widetilde{Z}\|_V^2 + 2\mathbb{E}[\|Z - \widetilde{Z}\|_V^2]$. Combined with Jensen's inequality and the first estimate, we conclude that

$$\mathbb{E}\Big[\exp\Big(\tfrac{\|Z-\mathbb{E}[Z]\|_V^2}{2^2\|Z-\widetilde{Z}\|_{L_\psi(\Omega;V)}^2}\Big)\Big] \le \mathbb{E}\Big[\exp\Big(\tfrac{\|Z-\widetilde{Z}\|_V^2}{2\|Z-\widetilde{Z}\|_{L_\psi(\Omega;V)}^2}\Big)\Big] \exp\Big(\tfrac{\mathbb{E}[\|Z-\widetilde{Z}\|_V^2]}{2\|Z-\widetilde{Z}\|_{L_\psi(\Omega;V)}^2}\Big) \le \mathrm{e}^{1/2}\mathrm{e}^{1/2} = \mathrm{e}.$$

□

We refer to $Z \in L_\psi(\Omega; V)$ as a V-valued random variable having *sub-Gaussian tail behavior*. Different notions of sub-Gaussianity are available in the literature, such as sub-Gaussianity w.r.t. an operator [9, Def. 1.1] or an orthonormal system [9, Def. 2.1], weak sub-Gaussianity [324, Def. 4.1], and γ-sub-Gaussianity [118]. Their relationships and properties are discussed, for example, in [118, 205]. Some of these notions imply sub-Gaussian tail behavior [118, Thms. 3.4 and 4.3]. We state common examples of V-valued random variables with sub-Gaussian tail behavior from the literature. Real-valued sub-Gaussian random variables, which are defined on p. ix, have sub-Gaussian tails.

Lemma 4.2.2. *If $\xi : \Omega \to \mathbb{R}$ is sub-Gaussian with parameter τ, then $\|\xi\|_{L_\psi(\Omega;\mathbb{R})}^2 \le 2\tau^2/(1 - \exp(-2))$.*

Proof. If $\tau = 0$, then $\xi = 0$ [57, p. 5]. Now, let $\tau > 0$. We have $\mathbb{E}[\exp(s\xi^2/(2\tau^2))] \le 1/(1-s)^{1/2}$ for all $s \in [0, 1)$ [57, Lem. 1.6 (p. 9)]. Choosing $s = 1 - \exp(-2) \in (0, 1)$ yields the claim. □

Lemma 4.2.3. *If H is a separable Hilbert space, and $Z : \Omega \to H$ is centered Gaussian, then $\|Z\|^2_{L_\psi(\Omega;H)} \le 2\mathbb{E}[\|Z\|^2_H]/(1-\exp(-2))$.*

Proof. If $\mathbb{E}[\|Z\|^2_H] = 0$, then $\|Z\|^2_{L_\psi(\Omega;H)} = 0$. Now, let $\sigma^2 = \mathbb{E}[\|Z\|^2_H] > 0$. Let $\xi : \Omega \to \mathbb{R}$ be a standard normal random variable. We define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = \exp(sx/(2\sigma^2))$, where $s = 1-\exp(-2)$. Since f is convex, [259, Rem. 4] ensures $\mathbb{E}[f(\|Z\|^2_H)] \le \mathbb{E}[f(\sigma^2\xi^2)]$. Combined with $\mathbb{E}[f(\sigma^2\xi^2)] = \mathbb{E}[\exp(s\xi^2/2)] = 1/(1-s)^{1/2}$ [57, p. 9], we obtain the claim. □

The following are further examples of V-valued random variables with sub-Gaussian tail behaviors: Rademacher series [345, p. 5], Gaussian random variables (due to the Landau–Shepp–Fernique theorem [345, p. 7], [356, Thm. 2.1.2]), certain Besov "priors" [87, Thm. 5], γ-sub-Gaussian random variables [118, Thm. 3.4], $L^p(\mathcal{D})$-valued sub-Gaussian vectors [118, Thm. 4.3], and sub-Gaussian random series [56, Thm. 1.10.3]. We provide examples of solutions to PDEs with random inputs that have sub-Gaussian tail behavior in section 4.5.1.

4.2.2 Uniformly Smooth and Quasi-Smooth Banach Spaces

We introduce a class of Banach spaces based on the notions used in [256, sect. 2], [166, Def. 2.1], [13, p. 468], and [345, Def. 3.1.2 and Prop. 3.1.2].

Definition 4.2.4. *Let $(V, \|\cdot\|_V)$ be a Banach space.*

(a) *The function $\|\cdot\|^2_V$ is* $(2,\kappa)$-smooth *if $\kappa \ge 1$ and*

$$\|x+y\|^2_V + \|x-y\|^2_V \le 2\|x\|^2_V + 2\kappa\|y\|^2_V \quad \textit{for all} \quad x, y \in V. \tag{4.2.2}$$

The space $(V, \|\cdot\|_V)$ is $(2,\kappa)$-smooth *if $\|\cdot\|^2_V$ is $(2,\kappa)$-smooth. We refer to $(V, \|\cdot\|_V)$ as* 2-uniformly smooth *if it is $(2,\kappa)$-smooth for some $\kappa \ge 1$.*

(b) *The space $(V, \|\cdot\|_V)$ is called* $(2,\kappa)$-quasi-smooth *if $\kappa \ge 1$ and there exists $\bar\kappa \in [1,\kappa]$ and a norm $|||\cdot|||_V$ on V such that $(V, |||\cdot|||_V)$ is $(2,\bar\kappa)$-smooth and*

$$\|x\|^2_V \le |||x|||^2_V \le (\kappa/\bar\kappa)\|x\|^2_V \quad \textit{for all} \quad x \in V. \tag{4.2.3}$$

We refer to $(V, \|\cdot\|_V)$ as 2-quasi-smooth *if it is $(2,\kappa)$-quasi-smooth for some $\kappa \ge 1$.*

In Definition 4.2.4, we restrict the smoothness constant κ to $[1,\infty)$ because whenever $(V, \|\cdot\|_V)$ is a Banach space with $V \ne \{0\}$ and (4.2.2) holds, then $\kappa \ge 1$. Our definition of a $(2,\kappa)$-smooth Banach space is equivalent to that in [166, 242] (see Lemma 4.7.6), and is compatible with that in [256, 257, 13]. (The constant D in [256, p. 1680] and [257, p. 55], and K in [13, p. 468] equal $\sqrt{\kappa}$.) The notion of a 2-uniformly smooth Banach space used here is equivalent to that used in the literature on the geometry in Banach spaces [345, Prop. 3.1.2], [262, Prop. 10.31].

The squared norm of a 2-uniformly smooth Banach space V is uniformly Fréchet differentiable [50, Prop. 4.2.14], that is, for $g = \|\cdot\|^2_V$ we have $(g(x+th)-g(x))/t \to \mathrm{D}g(x)[h]$ as $t \to 0^+$ uniformly for $h \in V$ with $\|h\|_V = 1$ and $x \in V$. Every 2-quasi-smooth Banach space is reflexive [233, Cor. 1.1] and has (Rademacher-)type 2 [212, Lem. 2.2].

The notion of $(2,\kappa)$-quasi-smoothness is introduced in [166, 242] as κ-*regularity*; see also [246, pp. 89–92]. The space $(\mathbb{R}^n, \|\cdot\|_\infty)$ is 2-quasi-smooth but not 2-uniformly smooth [166, Ex. 3.2]. Each finite-dimensional Banach space V is $(2,\dim(V))$-quasi-smooth [166, Ex. 3.1].

Every Hilbert space is $(2,1)$-smooth by the parallelogram identity. Moreover, if $(V, \|\cdot\|_V)$ is $(2,1)$-smooth, then $(V, \|\cdot\|_V)$ is a Hilbert space. To verify this assertion, it suffices to show that (4.2.2) implies the parallelogram identity [211, p. 53]. Substituting x by $(\bar x + \bar y)/2$ and y by $(\bar x - \bar y)/2$ in (4.2.2), we find that $\|\bar x\|^2_V + \|\bar y\|^2_V \le (1/2)\|\bar x + \bar y\|^2_V + (1/2)\|\bar x - \bar y\|^2_V$ for all $\bar x$, $\bar y \in V$. Combined with (4.2.2), we conclude that the parallelogram identity holds.

The notion of 2-quasi-smoothness provides flexibility over that of 2-uniform smoothness: (i) Some spaces are 2-quasi-smooth but fail to be 2-uniformly smooth, such as $(\mathbb{R}^n, \|\cdot\|_\infty)$. (ii) We apply MLMC mean estimators to approximate expected values of parameterized PDEs, which we discretize using finite elements. Multilevel corrections (see section 4.3) are then differences of random variables with values in the corresponding finite element spaces. For example, we show that the standard finite element spaces, defined by piecewise linear, continuous basis functions, are 2-quasi-smooth with the smoothness constant depending only logarithmically on the dimension of the finite element space (see section 4.4).
We provide further examples of 2-uniformly smooth and 2-quasi-smooth Banach spaces.

Lemma 4.2.5 ([166, Ex. 3.2], [242, Ex. 2.1])**.** *For $2 \leq p \leq \infty$, $(\mathbb{R}^n, \|\cdot\|_p)$ is $(2,\kappa)$-quasi-smooth with $\kappa = \inf_{2\leq\rho\leq p,\, \rho<\infty} \{ (\rho-1)n^{2/\rho-2/p} \}$. If $n \geq 3$, then $\kappa \leq (2\ln(n)-1)\mathrm{e}$.*

The bound on κ in Lemma 4.2.5 follows from [101, Thm. 2.2].

Theorem 4.2.6 ([256, Prop. 2.1], [101, Cor. 2.8], [262, Thm. 10.32], [348, Cor. 2], [200, Thm. 4.1], [13, Prop. 5])**.** *If $(\mathcal{O}, \mathcal{A}, \nu)$ is a σ-finite measurable space and $p \in [2,\infty)$, then $L^p(\mathcal{O}, \mathcal{A}, \nu; \mathbb{R})$ is $(2, p-1)$-smooth.*

Proposition 4.2.7. *If $s \in \mathbb{N}_0$, $2 \leq p < \infty$, and $\mathcal{D} \subset \mathbb{R}^d$ is a bounded domain, then $(W^{s,p}(\mathcal{D}), \|\cdot\|_{W^{s,p}(\mathcal{D})})$ is $(2, p-1)$-smooth.*

We prove Proposition 4.2.7 in section 4.7.1. The fact that the above Sobolev spaces are 2-uniformly smooth is known; see, e.g., [77, p. 54]. Proposition 4.2.7 provides the optimal 2-uniformly smoothness constant, $p-1$, for these Sobolev spaces. The optimality follows from the fact that the constant is optimal for $W^{0,p}(\mathcal{D}) = L^p(\mathcal{D})$ $(2 \leq p < \infty)$ [13, Prop. 3]. Further examples of 2-uniformly smooth and 2-quasi-smooth Banach spaces can be found in section 4.4 and in [242, 166, 13].
We present facts on 2-uniformly smooth and 2-quasi-smooth Banach spaces.

Lemma 4.2.8. *If $(V, \|\cdot\|_V)$ is $(2,\kappa)$-smooth and $F \subset V$ is a closed subspace, then $(W, \|\cdot\|_V)$ is $(2,\kappa)$-smooth. If $(V, \|\cdot\|_V)$ is $(2,\kappa)$-quasi-smooth and $W \subset V$ is a closed subspace, then $(W, \|\cdot\|_V)$ is $(2,\kappa)$-quasi-smooth.*

Lemma 4.2.9. *If $(W, \|\cdot\|_W)$ is a Banach space that is isometrically isomorphic to a $(2,\kappa)$-quasi-smooth Banach space, then $(W, \|\cdot\|_W)$ is $(2,\kappa)$-quasi-smooth.*

Proof. Let $(W, \|\cdot\|_W)$ be isometrically isomorphic to the $(2,\kappa)$-quasi-smooth space $(V, \|\cdot\|_V)$. Then, there exists a linear, bijective mapping $T : W \to V$ such that $\|Tx\|_V = \|x\|_W$ for all $x \in W$. Moreover, there exists $\bar{\kappa} \in [1,\kappa]$ and a norm $|||\cdot|||_V$ on V such that $(V, |||\cdot|||_V)$ is $(2,\bar{\kappa})$-smooth and (4.2.3) holds. We define the norm $|||\cdot|||_W = |||T\cdot|||_V$ on W. The $(2,\bar{\kappa})$-smoothness of $(V, |||\cdot|||_V)$ implies that of $(W, |||\cdot|||_W)$. Combined with (4.2.3), we conclude that $(W, \|\cdot\|_W)$ is $(2,\kappa)$-quasi-smooth. □

Lemma 4.2.10. *If $Z \in L^2(\Omega; V)$ and $\widetilde{Z} \in V$, then $\mathbb{E}[\|Z - \mathbb{E}[Z]\|_V^2] \leq 4\mathbb{E}[\|Z - \widetilde{Z}\|_V^2]$. If, in addition, $(V, \|\cdot\|_V)$ is $(2,\kappa)$-smooth, then $\mathbb{E}[\|Z - \mathbb{E}[Z]\|_V^2] \leq \kappa\mathbb{E}[\|Z - \widetilde{Z}\|_V^2] - \|\mathbb{E}[Z - \widetilde{Z}]\|_V^2$.*

Proof. We define $W = Z - \widetilde{Z}$. The bound $\mathbb{E}[\|W - \mathbb{E}[W]\|_V^2] \leq 4\mathbb{E}[\|W\|_V^2]$ follows from the triangle inequality and Jensen's inequality; see also [101, p. 148]. If $(V, \|\cdot\|_V)$ is $(2,\kappa)$-smooth, then Lemma 4.7.6 gives

$$\mathbb{E}[\|W - \mathbb{E}[W]\|_V^2] \leq \|\mathbb{E}[W]\|_V^2 - \mathbb{E}[\mathrm{D}g(\mathbb{E}[W])[W]] + \kappa\mathbb{E}[\|W\|_V^2],$$

where $g = \|\cdot\|_V^2$. Combined with $\mathrm{D}g(\mathbb{E}[W])[\mathbb{E}[W]] = 2\|\mathbb{E}[W]\|_V^2$ (see, e.g., [14, Ex. 2.32]) and $\mathbb{E}[\mathrm{D}g(\mathbb{E}[W])[W]] = \mathrm{D}g(\mathbb{E}[W])[\mathbb{E}[W]]$ (see, e.g., [159, eq. (1.2)]), we deduce the claim. □

4.2.3 Bounds on the Second Moment

We state bounds on the second moment for random sums taking values in 2-uniformly smooth and 2-quasi-smooth Banach spaces. Theorem 4.2.11 and Corollary 4.2.12 are essentially known from which we derive bounds on the second moment of random average. We use these bounds to analyze MLMC mean estimators. Proofs are presented in section 4.7.3.

Theorem 4.2.11. *Let $N \in \mathbb{N}$, and let $(V, \|\cdot\|_V)$ be a $(2,\kappa)$-smooth Banach space. Suppose that $(\xi_i)_{i\in\mathbb{N}_0} \subset L^2(\Omega;V)$ is a martingale-difference, and $x \in V$. Then*

$$\mathbb{E}\left[\left\|\sum_{i=1}^{N} \xi_i + x\right\|_V^2\right] \leq \|x\|_V^2 + \kappa \sum_{i=1}^{N} \mathbb{E}[\|\xi_i\|_V^2]. \tag{4.2.4}$$

In a slightly different form, Theorem 4.2.11 appears in [262, Thm. 10.22]. However, the bound (4.2.4) only depends on κ, rather than on an unspecified constant. For $x = 0$, versions of Theorem 4.2.11 are proven, for example, in [113, Thm. 2], [256, Prop. 2.5], [212, p. 155], [260, Prop. 2.4], and [166, p. 4]. The estimate (4.2.4) holds with equality and $\kappa = 1$ for independent, mean-zero Hilbert space-valued random variables [37, eq. (3.3)].
In order for the bound (4.2.4) to be valid for all $N \in \mathbb{N}$ and $x \in V$, and all independent, mean-zero V-valued random variables $\xi_1, \dots, \xi_N$, 2-uniform smoothness of $(V, \|\cdot\|_V)$ is a necessary and sufficient condition [345, p. 54].
Theorem 4.2.11 implies similar bounds for martingale-differences with values in 2-quasi-smooth Banach spaces.

Corollary 4.2.12. *Let $N \in \mathbb{N}$, and let $(V, \|\cdot\|_V)$ be a $(2,\kappa)$-quasi-smooth Banach space. Suppose that $(\xi_i)_{i\in\mathbb{N}_0} \subset L^2(\Omega;V)$ is a martingale-difference. Then*

$$\mathbb{E}\left[\left\|\sum_{i=1}^{N} \xi_i\right\|_V^2\right] \leq \kappa \sum_{i=1}^{N} \mathbb{E}[\|\xi_i\|_V^2]. \tag{4.2.5}$$

The inequality (4.2.5) is due to Juditsky and Nemirovski [166, p. 4] and Nemirovski [242, Prop. 3.1]. For $(\mathbb{R}^n, \|\cdot\|_r)$ with $2 \leq r \leq \infty$, Corollary 4.2.12 was first published by Nemirovski [240, Lem. 5.2.2]. The inequality (4.2.5) is known as *Nemirovski's inequality* [52, sect. 11.2], [101], [55, sect. 14.10.1].
Several approaches for deriving moment bounds for independent V-valued random variables are reviewed in [101]. For sums of independent mean-zero random variables with values in $(\mathbb{R}^n, \|\cdot\|_\infty)$, Nemirovski's inequality (4.2.4) yields tighter bounds than type-2-inequalities [101, pp. 147–149]. The fact that $(\mathbb{R}^n, \|\cdot\|_\infty)$ has type 2 has been used in [146, sect. 4], [197, pp. 1260–1261], and [135, eq. (2.8)] in the context of multilevel methods.
We refer the reader to [262, Chap. 12] for examples of interpolation spaces that are type 2 and nonreflexive [262, Rem. 12.1 and Cor. 12.19] and, hence, are not 2-quasi-smooth.
In order for the bound (4.2.5) to be valid for all $N \in \mathbb{N}$ and V-valued martingale-differences, 2-quasi-smoothness of $(V, \|\cdot\|_V)$ is a necessary and sufficient condition [262, Thm. 10.22 and Cor. 10.23].

Corollary 4.2.13. *Suppose that $\xi_{\ell,i} \in L^2(\Omega;V)$ for $i = 1, \dots, N_\ell$ and $\ell = 1, \dots, L$ are independent and mean-zero. If $(V, \|\cdot\|_V)$ is $(2,\kappa)$-quasi-smooth, then*

$$\mathbb{E}\left[\left\|\sum_{\ell=1}^{L} \frac{1}{N_\ell} \sum_{i=1}^{N_\ell} \xi_{\ell,i}\right\|_V^2\right] \leq \kappa \sum_{\ell=1}^{L} \frac{1}{N_\ell^2} \sum_{i=1}^{N_\ell} \mathbb{E}[\|\xi_{\ell,i}\|_V^2]. \tag{4.2.6}$$

If $(V, \|\cdot\|_V)$ is $(2,\kappa)$-smooth, then for each $x \in V$,

$$\mathbb{E}\Big[\Big\|\sum_{\ell=1}^{L}\frac{1}{N_\ell}\sum_{i=1}^{N_\ell}\xi_{\ell,i}+x\Big\|_V^2\Big] \le \|x\|_V^2+\kappa\sum_{\ell=1}^{L}\frac{1}{N_\ell^2}\sum_{i=1}^{N_\ell}\mathbb{E}[\|\xi_{\ell,i}\|_V^2]. \tag{4.2.7}$$

If the random variables in Corollary 4.2.13 take values in a (separable) Hilbert space, then equality holds in (4.2.7) with $\kappa = 1$ [37, Thm. 3.1].
We establish improvements over Corollary 4.2.12 for certain $W^{s,p}(\mathcal{D})$-valued random sums, where $\mathcal{D} \subset \mathbb{R}^d$ is a bounded domain. Let $p \ge 2$, $s \in \mathbb{N}_0$, $K \in \mathbb{N}$, and $\phi_k \in W^{s,p}(\mathcal{D})$ for $k = 1,\dots,K$. Further, suppose that $\xi_k : \Omega \to \mathbb{R}$ are sub-Gaussian random variables with parameter τ_k for $k = 1,\dots,K$. In this case, we define

$$T_K = \Big(\sum_{|\alpha|\le s}\|\sum_{k=1}^{K}\tau_k^2(\mathrm{D}^\alpha\phi_k)^2\|_{L^{p/2}(\mathcal{D})}^{p/2}\Big)^{1/p}. \tag{4.2.8}$$

Proposition 4.2.14. *Let $p \ge 2$, $s \in \mathbb{N}_0$ and $K \in \mathbb{N}$. Suppose that $\mathcal{D} \in \mathbb{R}^d$ is bounded and $\phi_k \in W^{s,p}(\mathcal{D})$ for $k = 1,\dots,K$. Let $\xi_k : \Omega \to \mathbb{R}$ be independent sub-Gaussian random variables with parameter $\tau_k > 0$ for $k = 1,\dots,K$. We define $Z = \sum_{k=1}^K \xi_k\phi_k$. Then $\mathbb{E}[\|Z\|_{W^{s,p}(\mathcal{D})}^p] \le 2(p/\mathrm{e})^{p/2}T_K^p$, where T_K is defined in (4.2.8). If, moreover, ξ_k are Gaussian with variance τ_k^2 for $k = 1,\dots,K$, then $\mathbb{E}[\|Z\|_{W^{s,p}(\mathcal{D})}^2] = \mathbb{E}[|\xi_1/\tau_1|^p]T_K^p$.*

Under the hypotheses of Proposition 4.2.14, the moment inequality given by Proposition 4.2.14 may be sharper than that of Corollary 4.2.12. Indeed, using the triangle inequality (applied to the norm $(\sum_{|\alpha|\le s}\|\cdot_\alpha\|_{L^{p/2}(\mathcal{D})}^{p/2})^{2/p}$ on the product space $\prod_{|\alpha|\le s}L^{p/2}(\mathcal{D})$ [1, Thm. 1.22]), the definition of the $\|\cdot\|_{W^{s,p}}$-norm (see p. viii), and (4.2.8), we obtain

$$T_K^2 = \Big[\sum_{|\alpha|\le s}\|\sum_{k=1}^{K}\tau_k^2(\mathrm{D}^\alpha\phi_k)^2\|_{L^{p/2}(\mathcal{D})}^{p/2}\Big]^{2/p} \le \sum_{k=1}^{K}\Big[\sum_{|\alpha|\le s}\|\tau_k^2(\mathrm{D}^\alpha\phi_k)^2\|_{L^{p/2}(\mathcal{D})}^{p/2}\Big]^{2/p} = \sum_{k=1}^{K}\tau_k^2\|\phi_k\|_{W^{s,p}}^2.$$

which is a loose bound in general. Moreover, we have $2^{2/p}p/\mathrm{e} < p-1$ for $p \ge 3$. On the other hand, the assumptions made in Proposition 4.2.14 are more restrictive than those of Corollary 4.2.12.

4.2.4 Exponential Tail Bounds

We state exponential bounds for the tail probabilities of certain sums of random averages, which are used for analyzing the tail behavior of the MLMC mean estimator in section 4.3. Proofs are presented in section 4.7.4.

Theorem 4.2.15. *Let $(V, \|\cdot\|_V)$ be a separable, $(2,\kappa)$-smooth Banach space, and $\xi_{\ell,i} \in L^1(\Omega;V)$ be independent and mean-zero such that $\mathbb{E}[\exp(\tau_\ell^{-2}\|\xi_{\ell,i}\|_V^2)] \le \mathrm{e}$ with $\tau_\ell > 0$ for i, $\ell = 1,2,\dots$. Then, for all $r \ge 0$, $L \in \mathbb{N}$, and every $N_\ell \in \mathbb{N}$, $\ell = 1,\dots,L$,*

$$\mathrm{Prob}\Big(\Big\|\sum_{\ell=1}^{L}\frac{1}{N_\ell}\sum_{i=1}^{N_\ell}\xi_{\ell,i}\Big\|_V \ge (\sqrt{\kappa}+r)\Big(\sum_{\ell=1}^{L}\frac{\tau_\ell^2}{N_\ell}\Big)^{1/2}\Big) \le \exp(-r^2/3). \tag{4.2.9}$$

If, in addition, $\|\xi_{\ell,i}\|_{L^\infty(\Omega;V)} \le \tau_\ell$ for i, $\ell = 1,2,\dots$, then the right-hand side in (4.2.9) improves to $\exp(-r^2/2)$.

The tail bounds in Theorem 4.2.15 can be improved when κ is small.

Theorem 4.2.16. *Let $(V, \|\cdot\|_V)$ be a separable, $(2,\kappa)$-smooth Banach space, and $\xi_{\ell,i} \in L^1(\Omega;V)$ be independent and mean-zero such that $\mathbb{E}[\exp(\tau_\ell^{-2}\|\xi_{\ell,i}\|_V^2)] \le \mathrm{e}$ with $\tau_\ell > 0$ for i, $\ell = 1,2,\ldots$. Then, for all $r \ge 0$, $L \in \mathbb{N}$, and every $N_\ell \in \mathbb{N}$, $\ell = 1,\ldots,L$,*

$$\mathrm{Prob}\biggl(\biggl\|\sum_{\ell=1}^{L}\frac{1}{N_\ell}\sum_{i=1}^{N_\ell}\xi_{\ell,i}\biggr\|_V \ge \sqrt{\kappa}r\biggl(\sum_{\ell=1}^{L}\frac{\tau_\ell^2}{N_\ell}\biggr)^{1/2}\biggr) \le 2\exp(-r^2/3). \tag{4.2.10}$$

If, in addition, $\|\xi_{\ell,i}\|_{L^\infty(\Omega;V)} \le \tau_\ell$ for i, $\ell = 1,2,\ldots$, then the right-hand side in (4.2.10) *improves to $2\exp(-r^2/2)$.*

Theorem 4.2.15 and the "renorming" lemma, Lemma 4.7.1, imply the following tail bounds.

Corollary 4.2.17. *Let $(V, \|\cdot\|_V)$ be a separable, $(2,\kappa)$-quasi-smooth Banach space, and $\xi_{\ell,i} \in L^1(\Omega;V)$ be independent and mean-zero such that $\mathbb{E}[\exp(\tau_\ell^{-2}\|\xi_{\ell,i}\|_V^2)] \le \mathrm{e}$ with $\tau_\ell > 0$ for i, $\ell = 1,2,\ldots$. Then, for all $r \ge 0$, $L \in \mathbb{N}$, and every $N_\ell \in \mathbb{N}$, $\ell = 1,\ldots,L$,*

$$\mathrm{Prob}\biggl(\biggl\|\sum_{\ell=1}^{L}\frac{1}{N_\ell}\sum_{i=1}^{N_\ell}\xi_{\ell,i}\biggr\|_V \ge (\sqrt{2\kappa}+\sqrt{2}r)\biggl(\sum_{\ell=1}^{L}\frac{\tau_\ell^2}{N_\ell}\biggr)^{1/2}\biggr) \le \exp(-r^2/3). \tag{4.2.11}$$

If, in addition, $\|\xi_{\ell,i}\|_{L^\infty(\Omega;V)} \le \tau_\ell$ for i, $\ell = 1,2,\ldots$, then the right-hand side in (4.2.11) *improves to $\exp(-r^2/2)$.*

Tail bounds similar to those in (4.2.9) and (4.2.11) are established for martingale-differences with values in (finite-dimensional) smooth spaces by Juditsky and Nemirovski [166, Thm. 4.1]. For $W^{s,p}(\mathcal{D})$-valued random sums, the tail bounds provided by Theorem 4.2.15 can be improved.

Theorem 4.2.18. *Let $p \ge 2$, $s \in \mathbb{N}_0$, and $K \in \mathbb{N}$. Suppose $\mathcal{D} \subset \mathbb{R}^d$ is a bounded domain and $\phi_k \in W^{s,p}(\mathcal{D})$ for $k = 1,\ldots,K$. Let ξ_k be independent mean-zero Gaussian random variables with variance $\tau_k^2 > 0$. Define $Z = \sum_{k=1}^K \xi_k\phi_k$ and $\gamma_p = (\mathbb{E}[|\xi_1/\tau_1|^p])^{1/p}$. Then $\gamma_p^p \le 2(p/\mathrm{e})^{p/2}$, and for each $\lambda \ge 0$, $\mathbb{E}[\cosh(\lambda\|Z\|_{W^{s,p}(\mathcal{D})})] \le \gamma_p^p\exp(\lambda^2T_K^2/2)$, where T_K is defined in* (4.2.8)*. In particular, for each $r > 0$,*

$$\mathrm{Prob}(\|Z\|_{W^{s,p}(\mathcal{D})} \ge r) \le 2\gamma_p^p\exp(-r^2/(2T_K^2)).$$

4.3 Multilevel Monte Carlo Mean Estimator

We introduce the MLMC mean estimator following [37, sect. 3] and [16, sect. 4.4].

(V) The space $(V, \|\cdot\|_V)$ is a Banach space, $V_1 \subset V_2 \subset \cdots \subset V$ are closed subspaces of V, and $(V_\ell, \|\cdot\|_V)$ is separable and $(2,\kappa_\ell)$-quasi-smooth for $\ell = 1,2,\ldots$. .

The choice $V_1 = V$ is possible in (V). However, in the literature on uncertainty quantification, the spaces (V_ℓ) are often finite element spaces [16, 15], which are finite-dimensional. The condition (V) allows us to estimate the expectation of continuous functions (see section 4.4).
Let $L \in \mathbb{N}$, and let the condition (V) hold. Let $X_\ell \in L^2(\Omega;V_\ell)$ for $\ell = 1,\ldots,L$. The MLMC mean estimator $\mathrm{E}^{\mathrm{ML}}[X_L]$ is defined by

$$\mathrm{E}^{\mathrm{ML}}[X_L] = \sum_{\ell=1}^{L}\mathrm{E}^{N_\ell}[Y_\ell], \qquad Y_\ell = X_\ell - X_{\ell-1}, \qquad X_0 = 0, \tag{4.3.1}$$

where the MC mean estimator $\mathrm{E}^{N_\ell}[Y_\ell]$ is defined by $\mathrm{E}^{N_\ell}[Y_\ell] = (1/N_\ell)\sum_{i=1}^{N} Y_{\ell,i}$. The functions Y_ℓ defined in (4.3.1) are called *multilevel corrections* [134, p. 6]. Here, $Y_{\ell,i}$ are independent for ℓ, $i = 1,2,\ldots$, and for each ℓ, $Y_{\ell,i}$ has the same probability distribution as Y_ℓ for $i = 1,2,\ldots$. For our analysis, we view $Y_{\ell,i}$ as random vectors defined on a common probability space; see, e.g., [44, pp. 148–149] for the standard construction of such a space.

4.3.1 Exponential Tail Bounds

The next theorem provides a cost analysis of the MLMC mean estimator using the exponential tail bounds established in Theorem 4.3.2. The cost of the MLMC method is defined by $\mathbb{E}[\mathrm{Cost}(\mathrm{E}^{\mathrm{ML}}[X_L])] = \sum_{\ell=1}^{L} N_\ell \mathbb{E}[\mathrm{Cost}(\mathrm{E}^{N_\ell}[Y_\ell])]$. The complexity model (4.3.2) is adapted from that developed by Giles [133, p. 609].
We recall that Y_ℓ are the multilevel corrections defined in (4.3.1), $\|\cdot\|_{L_\psi(\Omega;V)}$ is the Luxemburg norm defined in (4.2.1), $\lceil x \rceil_{\mathbb{N}} \in \mathbb{N}$ is the smallest number such that $x \leq \lceil x \rceil_{\mathbb{N}}$ for $x \in \mathbb{R}$, and $0 \notin \mathbb{N}$; see p. vii and section 4.2.

Theorem 4.3.1. *Let Assumption* (V) *hold, and* $(h_\ell) \subset \mathbb{R}_{++}$ *satisfy* $h_\ell = (1/s)h_{\ell-1}$ *for* $s \in \mathbb{N} \setminus \{1\}$. *Suppose there exist* α, β, $\gamma > 0$ *and* c_α, c_β, $c_\gamma > 0$ *such that*

$$\|\mathbb{E}[X_\ell] - \mathbb{E}[X]\|_V \leq c_\alpha h_\ell^\alpha, \quad \|Y_\ell - \mathbb{E}[Y_\ell]\|_{L_\psi(\Omega;V)}^2 \leq c_\beta h_\ell^\beta, \quad \text{and} \quad \mathbb{E}[\mathrm{Cost}(Y_\ell)] \leq c_\gamma h_\ell^{-\gamma}. \tag{4.3.2}$$

Then, for each $\varepsilon > 0$ *and* $\delta \in (0,1)$ *with* $L = \lceil (1/\alpha)\log_s(2c_\alpha h_1^\alpha/\varepsilon) + 1 \rceil_{\mathbb{N}}$, *there exist* $N_\ell \in \mathbb{N}$ $(\ell = 1, \ldots, L)$ *such that* $\mathrm{Prob}(\|\mathrm{E}^{\mathrm{ML}}[X_L] - \mathbb{E}[X]\|_V \geq \varepsilon) \leq \delta$, *and*

$$\mathbb{E}[\mathrm{Cost}(\mathrm{E}^{\mathrm{ML}}[X_L])] \lesssim \varepsilon^{-\gamma/\alpha} + (\sqrt{\kappa_L} + \sqrt{\ln(1/\delta)})^2 \begin{cases} \varepsilon^{-2} & \text{if} \quad \beta > \gamma, \\ \varepsilon^{-2}(\ln(\varepsilon^{-1}) + 1)^2 & \text{if} \quad \beta = \gamma, \\ \varepsilon^{-2-(\gamma-\beta)/\alpha} & \text{if} \quad \beta < \gamma. \end{cases} \tag{4.3.3}$$

We prove Theorem 4.3.1 using Theorem 4.3.2 and Lemma 4.3.3.

Theorem 4.3.2. *Let Assumption* (V) *hold. Suppose that* $L \in \mathbb{N}$ *and* $X \in L^2(\Omega; V)$. *For* $\ell = 1, \ldots, L$, *let* $X_\ell \in L^2(\Omega; V_\ell)$, *and let* $\tau_\ell > 0$ *fulfill* $\|Y_\ell - \mathbb{E}[Y_\ell]\|_{L_\psi(\Omega;V)} \leq \tau_\ell$. *Then, for all* $r > 0$,

$$\mathrm{Prob}\Big(\|\mathrm{E}^{\mathrm{ML}}[X_L] - \mathbb{E}[X]\|_V \geq (\sqrt{2\kappa_L} + \sqrt{2}r)\Big(\sum_{\ell=1}^{L} \tfrac{\tau_\ell^2}{N_\ell}\Big)^{1/2} + \|\mathbb{E}[X] - \mathbb{E}[X_L]\|_V\Big) \leq \mathrm{e}^{-\frac{r^2}{3}}, \tag{4.3.4}$$

where $\mathrm{E}^{\mathrm{ML}}[X_L]$ *and* Y_ℓ *are defined in* (4.3.1).

Proof. The definition of the MLMC mean estimator provided in (4.3.1) ensures that $(Y_{\ell,i})_{\ell,i}$ are independent. In particular, $\|Y_\ell - \mathbb{E}[Y_\ell]\|_{L_\psi(\Omega;V)} \leq \tau_\ell$ yields $\|Y_{\ell,i} - \mathbb{E}[Y_{\ell,i}]\|_{L_\psi(\Omega;V)} \leq \tau_\ell$ for $i = 1, \ldots, N_\ell$ and $\ell = 1, \ldots, L$. Lemma 4.2.8 implies that V_ℓ $(\ell = 1, \ldots, L-1)$ are $(2, \kappa_L)$-smooth. Moreover, we have

$$\mathrm{E}^{\mathrm{ML}}[X_L] - \mathbb{E}[X_L] = \sum_{\ell=1}^{L} \mathrm{E}^{N_\ell}[Y_\ell] - \mathbb{E}[Y_\ell] = \sum_{\ell=1}^{L} \sum_{i=1}^{N_\ell} \frac{Y_{\ell,i} - \mathbb{E}[Y_\ell]}{N_\ell}. \tag{4.3.5}$$

Using $\|\mathrm{E}^{\mathrm{ML}}[X_L] - \mathbb{E}[X]\|_V \leq \|\mathrm{E}^{\mathrm{ML}}[X_L] - \mathbb{E}[X_L]\|_V + \|\mathbb{E}[X_L] - \mathbb{E}[X]\|_V$, we obtain for all $\varepsilon \geq 0$,

$$\mathrm{Prob}\big(\|\mathrm{E}^{\mathrm{ML}}[X_L] - \mathbb{E}[X]\|_V \geq \varepsilon + \|\mathbb{E}[X_L] - \mathbb{E}[X]\|_V\big) \leq \mathrm{Prob}\big(\|\mathrm{E}^{\mathrm{ML}}[X_L] - \mathbb{E}[X_L]\|_V \geq \varepsilon\big).$$

Combined with (4.3.5), $\mathbb{E}[Y_{\ell,i}] = \mathbb{E}[Y_\ell]$, and Corollary 4.2.17, we obtain (4.3.4). □

The following lemma essentially follows from the proofs of [81, Thm. 1] and [37, Thm. 3.2] with the constants in (4.3.6) made explicit.

Lemma 4.3.3. *Let $(h_\ell) \subset \mathbb{R}_{++}$ satisfy $h_\ell = (1/s)h_{\ell-1}$ for $s \in \mathbb{N} \setminus \{1\}$. Suppose there exist α, β, γ c_α, $c_\beta > 0$ and (m_ℓ), $(\tau_\ell) \subset \mathbb{R}_+$ such that $m_\ell \le c_\alpha h_\ell^\alpha$ and $\tau_\ell^2 \le c_\beta h_\ell^\beta$ for $\ell = 1, 2, \ldots$. Then, for each $\epsilon > 0$, $\eta > 0$ with $L = \lceil \alpha^{-1} \log_s(c_\alpha \epsilon^{-1} h_1^\alpha) + 1 \rceil_{\mathbb{N}}$, there exists $N_\ell = N_\ell(\epsilon, \eta) \in \mathbb{N}$ $(\ell = 1, \ldots, L)$ such that*

$$N_\ell \le (1/s)^{(\beta+\gamma)/2} N_{\ell-1} + 1, \quad N_\ell \le (1/s)^{(\beta+\gamma)(\ell-1)/2} N_1 + 1, \quad m_L \le \epsilon, \quad \sum_{\ell=1}^{L} \tau_\ell^2 / N_\ell \le \epsilon^2 \eta.$$

Moreover, $L = 1$ if and only if $c_\alpha h_1^\alpha \le \epsilon$. If $L = 1$, then $N_1 h_1^{-\gamma} \le \lceil \epsilon^{-2} \eta^{-1} c_\beta h_1^\beta \rceil_{\mathbb{N}} h_1^{-\gamma}$, and otherwise

$$\begin{aligned} \sum_{\ell=1}^{L} N_\ell h_\ell^{-\gamma} \le{}& \frac{s^\gamma c_\alpha^{\gamma/\alpha}}{1 - s^{-\gamma}} h_1^{-\gamma} \cdot \epsilon^{-\gamma/\alpha} \\ &+ \eta^{-1} \begin{cases} \epsilon^{-2} \cdot c_\beta h_1^{\beta-\gamma} (1 - s^{-(\beta-\gamma)/2})^{-2} & \text{if} \quad \beta > \gamma, \\ \epsilon^{-2} \cdot c_\beta (\alpha^{-1} \log_s(c_\alpha \epsilon^{-1} h_1^\alpha) + 2)^2 & \text{if} \quad \beta = \gamma, \\ \epsilon^{-2-(\gamma-\beta)/\alpha} \cdot c_\beta (1 - s^{-(\gamma-\beta)/2})^{-2} s^{\gamma-\beta} c_\alpha^{(\gamma-\beta)/\alpha} & \text{if} \quad \beta < \gamma. \end{cases} \end{aligned} \tag{4.3.6}$$

Proof. The proof is presented in section 4.7.5. □

Proof of Theorem 4.3.1. We apply Theorem 4.3.2 and Lemma 4.3.3 to prove the claim. We define $\tau_\ell = \|Y_\ell - \mathbb{E}[Y_\ell]\|_{L_\psi(\Omega;V)}$ for $\ell = 1, \ldots, L$. To achieve $\mathrm{Prob}(\|\mathrm{E}^{\mathrm{ML}}[X_L] - \mathbb{E}[X]\|_V \ge \varepsilon) \le \delta$, we use the tail bound (4.3.4) established in Theorem 4.3.2. We require $\|\mathbb{E}[X] - \mathbb{E}[X_\ell]\|_V \le \varepsilon/2$, $\exp(-r^2/3) = \delta$, and $(\sqrt{2\kappa_L} + \sqrt{2}r)(\sum_{\ell=1}^{L} \tau_\ell^2/N_\ell)^{1/2} \le \varepsilon/2$. Since $r^2 = 3\ln(1/\delta)$, the latter requirement is fulfilled if

$$\sum_{\ell=1}^{L} \frac{\tau_\ell^2}{N_\ell} \le \frac{\varepsilon^2}{4} \frac{1}{(\sqrt{2\kappa_L} + \sqrt{6\ln(1/\delta)})^2}. \tag{4.3.7}$$

In order to apply Lemma 4.3.3, we use (4.3.2) and identify $m_\ell = \|\mathbb{E}[X] - \mathbb{E}[X_\ell]\|_V$, $\epsilon = \varepsilon/2$ and $\eta = 1/(\sqrt{2\kappa_L} + \sqrt{6\ln(1/\delta)})^2$. Now, the claim follows from Lemma 4.3.3. □

The proof of Theorem 4.3.1 also provides estimates for the number of levels L and the sample sizes N_ℓ $(\ell = 1, \ldots, L)$, and the rate $N_\ell \le (1/s)^{(\beta+\gamma)(\ell-1)/2} N_1 + 1$ through Lemma 4.3.3. The sample sizes only depend logarithmically on $1/\delta$ since $\eta^{-1} = (\sqrt{2\kappa_L} + \sqrt{6\ln(1/\delta)})^2$ (see the proof of Theorem 4.3.1).
The tail bound (4.3.4) can be improved, for example, when V is a Hilbert space using Theorem 4.2.16 instead of Corollary 4.2.17.
Theorem 4.3.2 implies exponential tail bounds for the MC estimator.

Corollary 4.3.4. *Let $(V, \|\cdot\|_V)$ be a Banach space, $V_L \subset V$, and $(V_L, \|\cdot\|_V)$ be a separable, $(2, \kappa_L)$-quasi-smooth Banach space, and $X \in L^2(\Omega; V)$, $X_L \in L^2(\Omega; V_L)$. Suppose $\sigma_L > 0$ satisfies $\|X_L - \mathbb{E}[X_L]\|_{L_\psi(\Omega;V)} \le \sigma_L$. Then, for all $r \ge 0$ and each $N \in \mathbb{N}$,*

$$\mathrm{Prob}\big(\|\mathrm{E}^N[X_L] - \mathbb{E}[X]\|_V \ge \tfrac{(\sqrt{2\kappa_L}+\sqrt{2}r)\sigma_L}{\sqrt{N}} + \|\mathbb{E}[X_L] - \mathbb{E}[X]\|_V\big) \le \exp(-r^2/3), \tag{4.3.8}$$

where E^N is the MC mean estimator.

Proof. Theorem 4.3.2 applied with $L = 1$ and $\tau_L = \sigma_L$ implies the claim. □

The following lemma provides bounds on the Luxemburg norm of $Y_\ell - \mathbb{E}[Y_\ell]$.

Lemma 4.3.5. *If $\ell \in \mathbb{N}$ and X, X_ℓ, $X_{\ell-1} \in L_\psi(\Omega; V)$, then $\|Y_\ell - \mathbb{E}[Y_\ell]\|_{L_\psi(\Omega;V)} \leq 2\|X_{\ell-1} - X\|_{L_\psi(\Omega;V)} + 2\|X_\ell - X\|_{L_\psi(\Omega;V)}$.*

Proof. Using the definition of Y_ℓ provided in (4.3.1) and the triangle inequality, we obtain

$$\|Y_\ell - \mathbb{E}[Y_\ell]\|_{L_\psi(\Omega;V)} \leq \|X_{\ell-1} - X - \mathbb{E}[X_{\ell-1} - X]\|_{L_\psi(\Omega;V)} + \|X_\ell - X - \mathbb{E}[X_\ell - X]\|_{L_\psi(\Omega;V)}.$$

Combining the estimate with Lemma 4.2.1 yields the assertion. □

The next lemma is used to verify the first two conditions in (4.3.2) for a class of linear elliptic PDEs in section 4.5.

Lemma 4.3.6. *Let $(h_\ell) \subset \mathbb{R}_{++}$ satisfy $h_\ell = (1/s)h_{\ell-1}$ for some $s \in \mathbb{N} \setminus \{1\}$. If $\|X - X_\ell\|_{L_\psi(\Omega;V)} \leq ch_\ell^t$ for $\ell = 1, 2, \ldots$, with $c > 0$ and $t > 0$, then the first two conditions in (4.3.2) hold with $\alpha = t$, $c_\alpha = c$, and $\beta = 2t$, $c_\beta = 4(1 + s^t)^2c^2$.*

Proof. Jensen's inequality and Lemma 4.2.1 yield $\|\mathbb{E}[X_\ell] - \mathbb{E}[X]\|_V^2 \leq \mathbb{E}[\|X_\ell - X\|_V^2] \leq \|X_\ell - \mathbb{E}[X_\ell]\|_{L_\psi(\Omega;V)}^2$. Hence, $\|\mathbb{E}[X_\ell] - \mathbb{E}[X]\|_V \leq ch_\ell^t$, $c_\alpha = c$, and $\alpha = t$. Lemma 4.3.5 gives $\|Y_\ell - \mathbb{E}[Y_\ell]\|_{L_\psi(\Omega;V)} \leq 2\|X_\ell - X\|_{L_\psi(\Omega;V)} + 2\|X_{\ell+1} - X\|_{L_\psi(\Omega;V)}$. Combined with $h_\ell = (1/s)h_{\ell-1}$, we obtain $\|Y_\ell - \mathbb{E}[Y_\ell]\|_{L_\psi(\Omega;V)} \leq 2(1 + s^t)ch_\ell^t$. Thus, $\beta = 2t$ and $c_\beta = 4(1 + s^t)^2c^2$. □

4.3.2 Sample Size Estimation and Cost Comparison

Sample size estimates are obtained via approximately minimizing the MLMC estimator's cost over a fixed variance by Giles [134, sect. 1.3]. We adapt this approach to our setting. For fixed $\varepsilon > 0$ and $\delta \in (0, 1)$, we compute N_ℓ ($\ell = 1, \ldots, L$) via approximately minimizing the estimator's cost $\sum_{\ell=1}^L N_\ell \mathbb{E}[\mathrm{Cost}(Y_\ell)]$ subject to (4.3.7). Replacing the variance of the MLMC mean estimator in [134, sect. 1.3] with $\tau_\ell^2 = \|Y_\ell - \mathbb{E}[Y_\ell]\|_{L_\psi(\Omega;V)}^2$ yields the estimate $N_\ell = c(\tau_\ell^2/\mathbb{E}[\mathrm{Cost}(Y_\ell)])^{1/2}$ for $\ell = 1, \ldots, L$, provided that $\mathbb{E}[\mathrm{Cost}(Y_\ell)] > 0$ and $\tau_\ell > 0$. Here $c = 4\varepsilon^{-2}(\sqrt{2\kappa_L} + \sqrt{6\ln(1/\delta)})^2 \sum_{\ell=1}^L (\tau_\ell^2 \mathbb{E}[\mathrm{Cost}(Y_\ell)])^{1/2}$.

We estimate the Luxemburg norm of $Y_\ell - \mathbb{E}[Y_\ell]$ for linear elliptic PDEs using stability and finite element error estimates in section 4.5. It may be difficult to obtain an accurate numerical estimate of the Luxemburg norm of $Y_\ell - \mathbb{E}[Y_\ell]$ via, for example, the MC approximation of the constraint in (4.2.1), as, for instance, $\exp(\|Y_\ell - \mathbb{E}[Y_\ell]\|_V^2/\tau^2)$ is generally heavy-tailed for $\tau > 0$; see also the discussions in [138, pp. 1036 and 1039]. We refer the reader to Wang and Ahmed [339] for an error analysis of stochastic programs with an expected value constraint.

We compare the costs of the MC estimator with that of the MLMC estimator for fixed accuracy $\varepsilon > 0$ and reliability $1 - \delta \in (0, 1)$. Since an upper bound on the cost of the MLMC estimator is provided by (4.3.3), it remains to compute the expected cost for the MC mean estimator.

For some (σ_ℓ), $(h_\ell) \subset \mathbb{R}_{++}$ and α, $\gamma > 0$, we assume that

$$\|\mathbb{E}[X_\ell] - \mathbb{E}[X]\|_V \lesssim h_\ell^\alpha, \quad \|X_\ell - \mathbb{E}[X_\ell]\|_{L_\psi(\Omega;V)} \leq \sigma_\ell, \quad \text{and} \quad \mathbb{E}[\mathrm{Cost}(X_\ell)] \lesssim h_\ell^{-\gamma}.$$

Under these conditions, $\mathbb{E}[\mathrm{Cost}(X_L)] \lesssim (\varepsilon/2)^{-\gamma/\alpha}$ and $h_L^\alpha \simeq \varepsilon$ yields $\|\mathbb{E}[X_L] - \mathbb{E}[X]\|_V \leq \varepsilon/2$. Moreover, $N \geq 4\sigma_L^2(\sqrt{2\kappa_L} + \sqrt{6\ln(1/\delta)})^2/\varepsilon^2$ and Corollary 4.3.4 ensure $\mathrm{Prob}(\|\mathrm{E}^N[X_L] - \mathbb{E}[X]\|_V \geq \varepsilon) \leq \delta$. Putting together the pieces, we conclude that

$$\mathbb{E}[\mathrm{Cost}(\mathrm{E}^N[X_L])] = N\mathbb{E}[\mathrm{Cost}(X_L)] \lesssim \sigma_L^2(\sqrt{\kappa_L} + \sqrt{\ln(1/\delta)})^2 \varepsilon^{-2-\gamma/\alpha}. \tag{4.3.9}$$

In order to show that the cost of the MLMC estimator is smaller than that of the sample mean, we require that the hypotheses of Theorem 4.3.1 hold, and $\mathbb{E}[\mathrm{Cost}(Y_\ell)] \simeq \mathbb{E}[\mathrm{Cost}(X_\ell)]$.

Furthermore, we assume that $\|X_L - \mathbb{E}[X_L]\|_{L_\psi(\Omega;V)} \simeq \|X - \mathbb{E}[X]\|_{L_\psi(\Omega;V)}$ and that the left-hand side in (4.3.9) is proportional to the right-hand side. When combined with (4.3.3), we obtain

$$\frac{\mathbb{E}[\mathrm{Cost}(\mathrm{E}^{\mathrm{ML}}[X_L])]}{\mathbb{E}[\mathrm{Cost}(\mathrm{E}^{N}[X_L])]} \lesssim (\sqrt{\kappa_L} + \sqrt{\ln(1/\delta)})^{-2}\varepsilon^2 + \begin{cases} \varepsilon^{\gamma/\alpha} & \text{if} \quad \beta > \gamma, \\ \varepsilon^{\gamma/\alpha}(\ln(\varepsilon^{-1}) + 1)^2 & \text{if} \quad \beta = \gamma, \\ \varepsilon^{\beta/\alpha} & \text{if} \quad \beta < \gamma. \end{cases}$$

Hence, for each $\delta \in (0,1)$, the cost of the MLMC estimator is smaller than that of the sample mean as $\varepsilon \to 0$.
The cost savings are similar to those obtained by Bierig and Chernov [37, pp. 587–588] for $\kappa_L = 1$, where the accuracy is measured using the mean-squared error. We note that the conditions $\mathbb{E}[\mathrm{Cost}(X_\ell)] \simeq \mathbb{E}[\mathrm{Cost}(Y_\ell)]$, $\|X_L - \mathbb{E}[X_L]\|_{L_\psi(\Omega;V)} \simeq \|X - \mathbb{E}[X]\|_{L_\psi(\Omega;V)}$ and $\mathbb{E}[\mathrm{Cost}(Y_\ell)] \le c_\gamma h_\ell^{-\gamma}$ are crucial for the cost comparison to be valid.[1]

4.4 Quasi-Smooth Approximations of Nonsmooth Banach Spaces

We show that certain finite-dimensional approximations of $(L^\infty(\mathcal{D}), \|\cdot\|_{L^\infty(\mathcal{D})})$ and of $(C(\bar{\mathcal{D}}), \|\cdot\|_{C(\bar{\mathcal{D}})})$ are $(2,\kappa)$-quasi-smooth with κ depending only logarithmically on the dimension of the subspaces. A canonical choice of such finite-dimensional subspaces are finite element spaces.

4.4.1 Space of Essentially Bounded Functions

The Banach space $(L^\infty(\mathcal{D}), \|\cdot\|_{L^\infty(\mathcal{D})})$ is nonreflexive [1, Thm. 2.35] and, hence, is not 2-quasi-smooth. We construct finite-dimensional, 2-quasi-smooth approximations of $L^\infty(\mathcal{D})$.
The following assumption is based on that used in [93, Assumption 3.3]. We choose $n_\ell \in \mathbb{N}$.

(A) $\varphi_k \in L^\infty(\mathcal{D})$, $x_k \in \mathcal{D}$, $\varphi_k(x_j) = \delta_{kj}$, $(k,\, j = 1,\dots,n_\ell)$, $\|\sum_{k=1}^{n_\ell} |\varphi_k|\|_{L^\infty(\mathcal{D})} = 1$.

Here, δ_{kj} is the Kronecker delta. Assumption (A) implicitly requires that the evaluations $\varphi_k(x_j)$ are well-defined. Under Assumption (A), we have $\|\varphi_k\|_{L^\infty(\mathcal{D})} = 1$ $(k = 1,\dots,n_\ell)$. The last condition in (A) implies that the supports of the functions φ_k are "almost" disjoint.
We define $V_\ell = \mathrm{span}\{\, \varphi_k : k = 1,\dots,n_\ell \,\}$. Assumption (A) ensures $\dim(V_\ell) = n_\ell$. Indeed, if $\alpha \in \mathbb{R}^{n_\ell}$ fulfills $\sum_{k=1}^{n_\ell} \alpha_k \varphi_k = 0$, then $\varphi_k(x_j) = \delta_{kj}$ implies $\alpha = 0$.

Example 4.4.1. We consider $\mathcal{D} = (0,1)^d$, discretize $\bar{\mathcal{D}}$ using a uniform grid with mesh width $h_\ell > 0$, and choose $n_\ell = 1/h_\ell^d$. For each cell, we define φ_k as the function that is equal to one on that cell and zero otherwise, and let x_k be the midpoint of the cell. Assumption (A) is fulfilled.

Assumption (A) also holds for continuous piecewise linear finite elements if $\mathcal{D} \subset \mathbb{R}^d$ is an interval for $d = 1$, a polygon for $d = 2$, or a polyhedron for $d = 3$ [54, Ex. 3.1.3], [93, Rem. 3.1].

Proposition 4.4.2. *Let Assumption* (A) *hold. Then* $(V_\ell, \|\cdot\|_{L^\infty(\mathcal{D})})$ *is* $(2,\kappa_\ell)$*-quasi-smooth with* $\kappa_\ell = \inf_{2\le r<\infty} \{\, (r-1)n_\ell^{2/r} \,\}$*, where* $n_\ell = \dim(V_\ell)$*. If* $n_\ell \ge 3$*, then* $\kappa_\ell \le (2\ln(n_\ell) - 1)\mathrm{e}$.

We apply Lemma 4.4.3 to prove Proposition 4.4.2.

[1] If these assumptions are violated, the MLMC mean estimator may have a larger variance than the sample mean. For example, consider $X = \xi$ and $X_\ell = \xi + |\xi|h_\ell$ with $\xi : \Omega \to \mathbb{R}$ being standard Gaussian. Then $|\mathbb{E}[X] - \mathbb{E}[X_\ell]| = (2/\pi)^{1/2}h_\ell$, and $Y_\ell - \mathbb{E}[Y_\ell] = (|\xi| - \mathbb{E}[|\xi|])(1+s)h_\ell \in L_\psi(\Omega;\mathbb{R})$. It is meaningful to choose $\mathbb{E}[\mathrm{Cost}(X_\ell)] = 1$. Since $Y_\ell = |\xi|(1+s)h_\ell$, we have $\mathbb{E}[\mathrm{Cost}(Y_\ell)] = 1$. Moreover, $\mathrm{Var}(\mathrm{E}^{\mathrm{ML}}[X_L]) = (h_1/N_1)\mathrm{Var}(|\xi|) + (1+s)\sum_{\ell=2}^{L}(h_\ell/N_\ell)\mathrm{Var}(|\xi|) > (h_L/N)\mathrm{Var}(|\xi|) = \mathrm{Var}(\mathrm{E}^{N}[X_L])$ for all $N \ge \min_\ell N_\ell$ and $L \in \mathbb{N}$. However $\mathbb{E}[\mathrm{Cost}(\mathrm{E}^{\mathrm{ML}}[X_L])] \le \mathbb{E}[\mathrm{Cost}(\mathrm{E}^{N}[X_L])]$ if and only if $N \ge \sum_{\ell=1}^{L} N_\ell$.

Lemma 4.4.3. *If Assumption* (A) *holds, then* $(V_\ell, \|\cdot\|_{L^\infty(\mathcal{D})})$ *is isometrically isomorphic to* $(\mathbb{R}^{n_\ell}, \|\cdot\|_\infty)$*, where* $n_\ell = \dim(V_\ell)$*.*

Proof. We define $T : V_\ell \to \mathbb{R}^{n_\ell}$ by $Tz = \alpha$ with $z = \sum_{k=1}^{n_\ell} \alpha_k \varphi_k$. The mapping T is linear and bijective. Fix $z \in V_\ell$. We deduce the existence of $\alpha \in \mathbb{R}^{n_\ell}$ with $z = \sum_{k=1}^{n_\ell} \alpha_k \varphi_k$. Using $\| \sum_{k=1}^{n_\ell} |\varphi_k| \|_{L^\infty(\mathcal{D})} = 1$, we have $\|z\|_{L^\infty(\mathcal{D})} \le \|\alpha\|_\infty \| \sum_{k=1}^{n_\ell} |\varphi_k| \|_{L^\infty(\mathcal{D})} = \|\alpha\|_\infty$. Now, fix $k \in \arg\max_{k=1,\dots,n_\ell} |\alpha_k|$. Since $\varphi_k(x_j) = \delta_{kj}$, we obtain $\|z\|_{L^\infty(\mathcal{D})} \ge |\alpha_k \varphi_k(x_k)| = \|\alpha\|_\infty$. Hence $\|Tz\|_\infty = \|z\|_{L^\infty(\mathcal{D})}$. □

Proof of Proposition 4.4.2. The space $(V_\ell, \|\cdot\|_{L^\infty(\mathcal{D})})$ is a closed, finite-dimensional subset of $L^\infty(\mathcal{D})$ and, hence, is separable Banach space. Now, the claim follows from Lemmas 4.2.5, 4.2.9 and 4.4.3. □

4.4.2 Space of Continuous Functions

We present a consequence of Proposition 4.4.2.

Corollary 4.4.4. *Let Assumption* (A) *hold and* $\varphi_k \in C(\bar{\mathcal{D}})$ *for* $k = 1, \dots, n_\ell$*. Then* $(V_\ell, \|\cdot\|_{C(\bar{\mathcal{D}})})$ *is* $(2, \kappa_\ell)$*-quasi-smooth with* $\kappa_\ell = \inf_{2 \le r < \infty} \{ (r-1) n_\ell^{2/r} \}$*, where* $n_\ell = \dim(V_\ell)$*.*

Proof. The assumptions ensure that $(V_\ell, \|\cdot\|_{C(\bar{\mathcal{D}})})$ is a closed, finite-dimensional subset of $(C(\bar{\mathcal{D}}), \|\cdot\|_{C(\bar{\mathcal{D}})})$ and, hence, it is a Banach space. Since $\|\cdot\|_{C(\bar{\mathcal{D}})} = \|\cdot\|_{L^\infty(\mathcal{D})}$ on V_ℓ, Lemma 4.4.3 and Proposition 4.4.2 yield the claim. □

4.5 Application to Linear Elliptic PDEs with Random Inputs

We consider a class of linear elliptic PDEs with random inputs and provide conditions that allow the application of Theorem 4.3.1 on the complexity of the MLMC mean estimator. It turns out that the solution of a linear elliptic PDE has sub-Gaussian tail behavior if the random diffusion coefficient is uniformly bounded, and the right-hand side has sub-Gaussian tail behavior. We provide an example of a elliptic PDE with a log-normal diffusion coefficient such that its solution is "heavy-tailed."

Our error analysis is restricted to discretization errors, resulting from a finite element approximation; see, e.g., [16, 70]. A complete discussion may further include the analysis of truncation errors of data defined by random series (see, e.g., [69, sect. 4], [311, sect. 4.1]), and of quadrature errors (see, e.g., [70, sect. 3.3]).

We consider the linear elliptic PDE with random inputs: Find $y : \Omega \to H_0^1(\mathcal{D})$ such that, w.p. 1,

$$(\kappa(\omega)\nabla y(\omega), \nabla v)_{L^2(\mathcal{D})^d} = (b(\omega), v)_{L^2(\mathcal{D})} \quad \text{for all} \quad v \in H_0^1(\mathcal{D}), \tag{4.5.1}$$

where $\kappa : \Omega \to L^\infty(\mathcal{D})$ is the random diffusion coefficient and $b : \Omega \to L^2(\mathcal{D})$. We define

$$\kappa_{\min}(\omega) = \operatorname{ess\,inf}_{x \in \mathcal{D}} \kappa(\omega)(x), \quad \text{and} \quad \kappa_{\max}(\omega) = \operatorname{ess\,sup}_{x \in \mathcal{D}} \kappa(\omega)(x). \tag{4.5.2}$$

The following conditions on (4.5.1) are based on those used in [70, sect. 2.2], [311, sect. 2.1].

(D1) The bounded domain $\mathcal{D} \subset \mathbb{R}^d$ has a $C^{0,2}$-boundary.[2]
(D2) $\kappa \in L^p(\Omega; C^1(\bar{\mathcal{D}}))$ for all $p \in [1, \infty)$.
(D3) $\kappa_{\min} > 0$, and $1/\kappa_{\min} \in L^p(\mathcal{D})$ for all $p \in [1, \infty)$.
(D4) $b \in L^{p^*}(\Omega; L^2(\mathcal{D}))$ for some $p^* \in [2, \infty)$.

[2] We refer the reader to [151, Def. 1.13] for the definition of a $C^{0,2}$-boundary.

Assumptions (D2) and (D3) ensure the measurability of the functions $\kappa_{\min}$ and $\kappa_{\max}$.

Lemma 4.5.1. *If Assumptions* (D1)–(D4) *hold, then the PDE* (4.5.1) *has a unique solution* $y : \Omega \to H_0^1(\mathcal{D})$*, w.p.* 1*,* $|y(\omega)|_{H^1(\mathcal{D})} \leq (C_{\mathcal{D}}/\kappa_{\min}(\omega))\|b(\omega)\|_{L^2(\mathcal{D})}$*, and* $y \in L^p(\Omega; H_0^1(\mathcal{D}))$ *for all* $1 \leq p < p^*$*, where* $C_{\mathcal{D}} = \sup_{v \in H_0^1(\mathcal{D})\setminus\{0\}} \{ \|v\|_{L^2(\mathcal{D})}/|v|_{H^1(\mathcal{D})} \}$ *is Friedrichs' constant of* $\mathcal{D}$*.*

Proof. Owing to (D1)–(D4), [12, Lem. 2.1] yields the claim (see also [70, Thm. 2.2]). □

Throughout the section, let $y \in L^p(\Omega; H_0^1(\mathcal{D}))$ be the solution to (4.5.1). We define the multilevel corrections through solutions of finite element approximations of the PDE (4.5.1). In order to avoid introducing triangulations and finite element spaces, we impose the following condition.

(D5) There exists a sequence of nested subspaces $(Y_h) \subset H_0^1(\mathcal{D})$ such that $\dim(Y_h) \lesssim (1/h)^d$ and $\inf_{v_h \in Y_h} |v - v_h|_{H^1(\mathcal{D})} \lesssim |v|_{H^2(\mathcal{D})} h$ for all $h > 0$ and $v \in H^2(\mathcal{D}) \cap H_0^1(\mathcal{D})$.

According to [70, Lem. 3.7] and Friedrichs' inequality (see, e.g., [151, Thm. 1.13]), Assumption (D5) holds if, for each $h > 0$, Y_h is the space of continuous, piecewise linear functions defined on a polygonal approximation of $\mathcal{D}$, and $\mathcal{D}$ satisfies (D1) and [70, Assumption A4].
For $h > 0$, we consider the approximation of the PDE (4.5.1): Find $y_h : \Omega \to Y_h$ such that, w.p. 1,

$$(\kappa(\omega)\nabla y_h(\omega), \nabla v_h)_{L^2(\mathcal{D})^d} = (b(\omega), v_h)_{L^2(\mathcal{D})} \quad \text{for all} \quad v \in Y_h. \tag{4.5.3}$$

We define the random variables C_1, $C_2 : \Omega \to \mathbb{R}_{++}$ by[3]

$$C_1(\omega) = \frac{\kappa_{\max}(\omega)\|\kappa(\omega)\|_{C^1(\bar{\mathcal{D}})}}{\kappa_{\min}(\omega)^3}, \quad \text{and} \quad C_2(\omega) = \left(\frac{\kappa_{\max}(\omega)}{\kappa_{\min}(\omega)}\right)^{1/2}. \tag{4.5.4}$$

Lemma 4.5.2. *If Assumptions* (D1)–(D5) *hold, then for each* $h > 0$*, the discretized PDE* (4.5.3) *has a unique solution* $y_h : \Omega \to Y_h$*,* $y_h \in L^p(\Omega; H_0^1(\mathcal{D}))$ *for all* $1 \leq p < p^*$*, and w.p.* 1*,*

$$|y(\omega) - y_h(\omega)|_{H^1(\mathcal{D})} \lesssim C_1(\omega)C_2(\omega)\|b(\omega)\|_{L^2(\mathcal{D})} h \quad \textit{for all} \quad h > 0. \tag{4.5.5}$$

Proof. The existence and uniqueness of $y_h : \Omega \to Y_h$ and $y_h \in L^p(\Omega; H_0^1(\mathcal{D}))$ can be established using similar arguments as in the proof of [70, Thm. 2.2]. The error estimate (4.5.5) essentially follows from the proof of [70, Thm. 3.9]. We have $\|y(\omega)\|_{H^2(\mathcal{D})} \lesssim C_1(\omega)\|b(\omega)\|_{L^2(\mathcal{D})}$ [70, Prop. 3.1]. Owing to (D3)–(D5), Céa's lemma yields $|y(\omega) - y_h(\omega)|_{H^1(\mathcal{D})} \leq C_2(\omega) \inf_{v_h \in Y_h} |y(\omega) - v_h|_{H^1(\mathcal{D})}$ [54, Rem. 2.8.5]. Combined with (D5), we obtain (4.5.5). □

Throughout the section, let $y_h \in L^p(\Omega; H_0^1(\mathcal{D}))$ be the solution to (4.5.3) for $h > 0$. The finite element error estimate (4.5.5) remains valid for domains other than those given by (D1), such as polygonal/polyhedral ones [311, p. 574], when replacing the random variable $C_1(\omega)$ defined in (4.5.4) with a different one [311, Lem. 5.2].

4.5.1 Light-Tailed Solutions

In order to show that the solution to (4.5.1) has sub-Gaussian tail behavior, we use the stability estimate from Lemma 4.5.1.

Lemma 4.5.3. *Let Assumptions* (D1)–(D5) *hold, and define* $\kappa_{\min}^* = \operatorname{ess\,inf}_{\omega \in \Omega} \kappa_{\min}(\omega)$*. Suppose that* $b \in L_\psi(\Omega; L^2(\mathcal{D}))$ *and* $\kappa_{\min}^* > 0$*. Then* $\|y\|_{L_\psi(\Omega; H_0^1(\mathcal{D}))} \leq (C_{\mathcal{D}}/\kappa_{\min}^*)\|b\|_{L_\psi(\Omega; L^2(\mathcal{D}))}$*.*

[3] The space $C^1(\bar{\mathcal{D}})$ is equipped with the norm $\|\cdot\|_{C^1(\bar{\mathcal{D}})}$ defined by $\|v\|_{C^1(\bar{\mathcal{D}})} = \sum_{|\alpha| \leq 1} \sup_{x \in \bar{\mathcal{D}}} |\mathrm{D}^\alpha v(x)|$.

Proof. Lemma 4.5.1 yields $|y(\omega)|_{H^1(\mathcal{D})} \leq (C_{\mathcal{D}}/\kappa^*_{\min})\|b(\omega)\|_{L^2(\mathcal{D})}$. Combined with the definition of the Luxemburg norm (see (4.2.1)), we obtain the claim. □

Lemma 4.5.4. *Under the conditions of Lemma 4.5.3 and $\kappa \in L^\infty(\Omega; C^1(\bar{\mathcal{D}}))$, we have $C_3^* = \operatorname{ess\,sup}_{\omega\in\Omega}\{\, C_1(\omega)C_2(\omega)\,\} < \infty$, and $\|y - y_h\|_{L_\psi(\Omega;H_0^1(\mathcal{D}))} \lesssim C_3^*\|b\|_{L_\psi(\Omega;L^2(\mathcal{D}))}h$.*

Proof. The claims follow from an application of Lemma 4.5.2. □

Under the hypotheses of Lemma 4.5.4, Lemmas 4.3.6 and 4.5.4 imply that the first two conditions in (4.3.2) are satisfied. The assumption on the random diffusion coefficient imposed by Lemma 4.5.4 are similar to those used by Barth, Schwab, and Zollinger [16, Prop. 3.6], with the difference being the requirement $\kappa \in L^\infty(\Omega; C^1(\bar{\mathcal{D}}))$ instead of $\kappa \in L^\infty(\Omega; W^{1,\infty}(\mathcal{D}))$.
We adapt an observation by Bharucha-Reid [36, p. 126] on the expectations of solutions to random linear operator equations to our setting. If κ and b are independent, then we can compute the expectation of $y : \Omega \to H_0^1(\mathcal{D})$ by that of the solution to the PDE (4.5.1) with b replaced by its mean $\mathbb{E}[b]$.[4] In particular, the right-hand side b is not required to have sub-Gaussian tails in order to apply our framework. Babuška, Nobile, and Tempone [12, Rem. 1] provide a motivation for why it may be reasonable to model κ and b as independent.

Lemma 4.5.5. *If Assumptions* (D1)–(D4) *hold, κ and b are independent, and $\bar{y} : \Omega \to H_0^1(\mathcal{D})$ solves $(\kappa(\omega)\nabla\bar{y}(\omega), \nabla v)_{L^2(\mathcal{D})^d} = (\mathbb{E}[b], v)_{L^2(\mathcal{D})}$ for all $v \in H_0^1(\mathcal{D})$, then $\mathbb{E}[y] = \mathbb{E}[\bar{y}]$.*

Proof. We define $A : \Omega \to \mathscr{L}(H_0^1(\mathcal{D}), H_0^1(\mathcal{D})^*)$ and $B : \Omega \to H_0^1(\mathcal{D})^*$ by

$$\langle A(\omega)y, v\rangle_{H_0^1(\mathcal{D})^*,H_0^1(\mathcal{D})} = (\kappa(\omega)\nabla y, \nabla v)_{L^2(\mathcal{D})^d} \quad \text{and} \quad \langle B(\omega), v\rangle_{H_0^1(\mathcal{D})^*,H_0^1(\mathcal{D})} = (b(\omega), v)_{L^2(\mathcal{D})}.$$

Both functions are well-defined by the Lax–Milgram lemma [151, Lem. 1.8], B is Bochner integrable by (D1) and (D4), and $\langle \mathbb{E}[B], v\rangle_{H_0^1(\mathcal{D})^*,H_0^1(\mathcal{D})} = (\mathbb{E}[b], v)_{L^2(\mathcal{D})}$ for all $v \in H_0^1(\mathcal{D})$ [36, p. 78]. Hence, the PDE (4.5.1) can be equivalently written as $A(\omega)y(\omega) = B(\omega)$, and the above PDE with deterministic right-hand side as $A(\omega)\bar{y}(\omega) = \mathbb{E}[B]$. Below, we establish the identities $\mathbb{E}[y] = \mathbb{E}[A^{-1}B] = \mathbb{E}[A^{-1}]\mathbb{E}[B] = \mathbb{E}[A^{-1}\mathbb{E}[B]] = \mathbb{E}[\bar{y}]$.
We show that A and A^{-1} are strongly measurable w.r.t. the uniform operator topology. Owing to (D2), κ is strongly measurable. Lemma 3.2.24 ensures that the mapping $\phi : C^1(\bar{\mathcal{D}}) \to \mathscr{L}(H_0^1(\mathcal{D}), H_0^1(\mathcal{D})^*)$ defined by $\langle \phi(\kappa)y, v\rangle_{H_0^1(\mathcal{D})^*,H_0^1(\mathcal{D})} = (\kappa\nabla y, \nabla v)_{L^2(\mathcal{D})^d}$ is (Lipschitz) continuous. Hence, $A = \phi \circ \kappa$ is strongly measurable [159, Cor. 1.1.11]. Since A^{-1} is the composition of a continuous function with A, it is also strongly measurable [159, Cor. 1.1.11].
Using (D3) and [151, Lem. 1.8], we obtain $\|A^{-1}\|_{\mathscr{L}(H_0^1(\mathcal{D})^*,H_0^1(\mathcal{D}))} \leq 1/\kappa_{\min} \in L^1(\mathcal{D})$ implying that A^{-1} is Bochner integrable. The independence of κ and b ensure that of A and B [44, pp. 398–399]. Hence, A^{-1} and B are independent [44, pp. 398–399]. Since B is Bochner integrable, we have $\mathbb{E}[B] \in H_0^1(\mathcal{D})^*$ [159, p. 13]. Putting together the pieces, we find that $\mathbb{E}[y] = \mathbb{E}[A^{-1}B] = \mathbb{E}[A^{-1}]\mathbb{E}[B]$ [160, Prop. 6.1.3]. The Bochner integrability of A^{-1} also ensures $\mathbb{E}[A^{-1}]\mathbb{E}[B] = \mathbb{E}[A^{-1}\mathbb{E}[B]]$ [36, p. 78]. Hence, $\mathbb{E}[y] = \mathbb{E}[\bar{y}]$. □

4.5.2 Heavy-Tailed Solutions

The solutions to linear elliptic PDEs with a log-normal diffusion coefficient may be "heavy-tailed," which we show on a model problem.

[4] If $\mathbb{E}[b]$ is not available in closed form, we may replace it by its sample mean. The approximation error can be studied with [221, Thm. 9.31].

Example 4.5.6. We define $\mathcal{D} = (0,1)$, $b = 2$ and $\kappa(\omega) = \exp(-\xi(\omega))$, where $\xi : \Omega \to \mathbb{R}$ is a Gaussian random variable with zero mean and unit variance. This simple problem defines an elliptic PDE with a log-normal diffusion coefficient [311, sect. 2], [12, sect. 1], [69, sect. 2].
The unique solution y to (4.5.1) is $y(\omega)(x) = \exp(\xi(\omega))x(1-x)$. We have $|y|_{H^1(\mathcal{D})} = \exp(\xi)/\sqrt{3}$. Since $\mathbb{E}[\exp(s\xi^2/2)] = 1/(1-s)^{1/2}$ if $s \in [0,1)$ and $\mathbb{E}[\exp(s\xi^2/2)] = \infty$ if $s \geq 1$ [57, p. 9], we have $\|y\|_{L_\psi(\Omega;H_0^1(\mathcal{D}))} = \infty$.
The distribution of $\exp(\xi)$ is referred to as *(moderately) heavy-tailed* in the literature [105, pp. 9 and 138], [114, p. 39]. Since $|y|_{H^1(\mathcal{D})} = \exp(\xi)/\sqrt{3}$, the tail probabilities of the random variable $|y|_{H^1(\mathcal{D})}$ are determined by those of $\exp(\xi)$. For all $r > 0$, we have

$$\frac{\exp(-r^2/2)}{\sqrt{\pi}(\sqrt{2 + r^2/2} + r/\sqrt{2})} \leq \mathrm{Prob}(\xi \geq r) \leq \frac{\exp(-r^2/2)}{\sqrt{\pi}(\sqrt{1 + r^2/2} + r/\sqrt{2})}, \tag{4.5.6}$$

[57, eq. (3.6), p. 227]. Hence, the tails of $\exp(\xi)$ decrease at a much slower rate than those of ξ.

Let $g \in L^0(\Omega; C(\bar{\mathcal{D}}))$ be a mean-zero, nondegenerate Gaussian random variable, and define the log-normal diffusion coefficient κ by $\kappa = \exp(g)$. Using (4.5.2), we have, w.p. 1, $1/\kappa_{\min}(\omega) \leq \exp(\|g(\omega)\|_{C(\bar{\mathcal{D}})})$. This estimate is typically used in combination with a stability estimate to deduce the existence of all higher-order moments of the solution to (4.5.1); see, e.g., [69, p. 218–219], [70, p. 325–326]. This estimate may not be used to deduce sub-Gaussian tail behavior of the solution to (4.5.1) since $\|\exp(\|g\|_{C(\bar{\mathcal{D}})})\|_{L_\psi(\Omega;\mathbb{R})} = \infty$. Indeed, we have $\mathbb{E}[\exp(\tau\|g\|^2_{C(\bar{\mathcal{D}})})] = \infty$ for all $\tau > 1/(2\sigma^2)$, where $\sigma^2 = \sup_{f \in C(\bar{\mathcal{D}})^*} \mathbb{E}[\langle f, g\rangle^2_{C(\bar{\mathcal{D}})^*, C(\bar{\mathcal{D}})}] > 0$ [356, Rem. 2.1.3].

4.5.3 Numerical Simulations

We perform simulations for the Hilbert space $(V, \|\cdot\|_V) = (H_0^1(\mathcal{D}), |\cdot|_{H^1(\mathcal{D})})$, $X = y$, and $X_\ell = y_{h_\ell}$, and an adaption of the model problem considered by Barth, Schwab, and Zollinger [16, sect. 6.2].

Implementation Details

We present implementation details for a practical version of the MLMC mean estimator which tries to ensure that the estimate is close to the true mean with high probability. The pseudo-code of the MLMC method, Algorithm 3, is as in [134, Alg. 1] with the differences that the squared Luxemburg norm of the multilevel corrections is estimated rather than their variance, and the number of samples is determined using a formula that depends both on the accuracy and the reliability. Our implementation is based on that of `pymlmc` [108] and the approaches developed by Giles [134, sect. 3]. We focus on discussing the main differences to the approaches used in [134, sect. 3].
We adapt the approach implemented in `pymlmc` [108] to adaptively estimate the sample sizes N_ℓ. Let $\varepsilon > 0$ be the accuracy, and let $1 - \delta \in (0,1)$ be the reliability. Motivated by the derivations in section 4.3.1 (cf. [134, eq. (3.1)]), we choose

$$N_\ell = \left\lceil (4/3)\varepsilon^{-2} \ln(1/\delta)(\tau_\ell^2/C_\ell)^{1/2} \sum_{\ell=1}^{L} (\tau_\ell^2 C_\ell)^{1/2} \right\rceil_{\mathbb{N}}, \tag{4.5.7}$$

where C_ℓ is an estimate of $\mathbb{E}[\mathrm{Cost}(Y_\ell)]$, and τ_ℓ is an approximation of $\|Y_\ell - \mathbb{E}[Y_\ell]\|_{L_\psi(\Omega;V)}$. The cost C_ℓ is either the average simulation time of $Y_{\ell,i}$ $(i = 1, \ldots, N_\ell)$ or the user-defined cost as in `pymlmc` [108]. Here, Y_ℓ are the multilevel corrections defined in (4.3.1) and $\|\cdot\|_{L_\psi(\Omega;V)}$ is the Luxemburg norm defined in (4.2.1).

Algorithm 3 Multilevel Monte Carlo Method (`ProbMLMC`)

Choose accuracy $\varepsilon > 0$, reliability $1 - \delta \in (0, 1)$, and initial sample size N_ℓ for $\ell = 1, 2$.
For $L = 3, 4, \ldots$

1. Evaluate (remaining) samples on the levels $1, \ldots, L$.
2. Estimate Luxemburg norm of $Y_\ell - \widetilde{Y}_\ell$ and estimate cost $\mathbb{E}[\mathrm{Cost}(Y_\ell)]$ $(\ell = 1, \ldots, L)$.
3. Compute sample size N_ℓ using (4.5.7) $(\ell = 1, \ldots, L)$.
4. Test for convergence.
5. Initialize sample size N_{L+1}.

We estimate the Luxemburg norm of $Y_\ell - \mathbb{E}[Y_\ell]$ by that of $Y_\ell - \widetilde{Y}_\ell$, where $\widetilde{Y}_\ell = \widetilde{X}_\ell - \widetilde{X}_{\ell-1}$ and $\widetilde{X}_\ell$ is the finite element solution on V_ℓ of the PDE with diffusion coefficient $\mathbb{E}[\kappa]$ and right-hand side $\mathbb{E}[b]$. Lemma 4.2.1 ensures $\|Y_\ell - \mathbb{E}[Y_\ell]\|_{L_\psi(\Omega;V)} \leq 2\|Y_\ell - \widetilde{Y}_\ell\|_{L_\psi(\Omega;V)}$. This choice does not require us to save the samples $Y_{\ell,i}$ $(i = 1, \ldots, N_\ell, \ell = 1, \ldots, L)$, and ensures $\|Y_\ell - \widetilde{Y}_\ell\|_{L_\psi(\Omega;V)} = 0$ if Y_ℓ is deterministic. If Y_ℓ is real-valued, it may be feasible to save the samples $Y_{\ell,i}$, which then allow for the computation of $|Y_{\ell,i} - \mathrm{E}^{N_\ell}[Y_\ell]|$. In contrast to computing of the Luxemburg norm, the sample variance of $Z : \Omega \to \mathbb{R}$ can be evaluated without using $\mathrm{E}^{N_\ell}[Z]$ with the one-pass method [68, eq. (1.3)].
For the random variable $Z = \|Y_\ell - \widetilde{Y}_\ell\|_V$ with samples $Z_1, \ldots, Z_{N_\ell}$, we estimate $\|Z\|_{L_\psi(\Omega;\mathbb{R})}$ via the solution to: Find $\tau > 0$ with $\mathrm{E}^{N_\ell}[\exp(|Z|^2/\tau^2)] - \mathrm{e} = 0$. When $Z_i \neq 0$ for some $i \in \{1, \ldots, N_\ell\}$, the root exists and is the Luxemburg norm of the sample $Z_1, \ldots, Z_{N_\ell}$; see section 4.2.1. We apply the implementation of Brent's method provided by `SciPy` [327] with initial values $\mathrm{E}^N[|Z|^2]^{1/2}$ and $\max_{i \in \{1,\ldots,N_\ell\}} |Z_i|$, which ensure that the root function has opposite signs; see section 4.2.1. As discussed in section 4.3.2, this estimate may provide an inaccurate approximation to $\|Z\|_{L_\psi(\Omega;\mathbb{R})}$. However, it allows us to adaptively estimate the constants in (4.3.2) by adapting the approaches developed by Giles [134, sect. 3].
We terminate Algorithm 3 if $\max_{\ell \in \{L-2, L-1, L\} \cap \mathbb{N}} \{ \|\mathrm{E}^{N_\ell}[Y_\ell]\|_V / s^{\alpha(L-\ell)} \} \leq (s^\alpha - 1)\varepsilon/2$ as in [134, pp. 279 and 284]. We provide a short motivation for this rule. We assume $\mathbb{E}[X_\ell] \to \mathbb{E}[X]$ as $\ell \to \infty$ and $\|\mathbb{E}[Y_\ell]\|_V \leq (1/s)^\alpha \|\mathbb{E}[Y_{\ell-1}]\|_V$ for some $\alpha > 0$, $s \in \mathbb{N} \setminus \{1\}$, and $\ell \geq 2$. Combined with the telescoping sum and the Cauchy–Schwarz inequality, we find that $\|\mathbb{E}[X_\ell] - \mathbb{E}[X]\|_V \leq \sum_{k=\ell}^\infty \|\mathbb{E}[Y_{k+1}]\|_V \leq \|\mathbb{E}[Y_{\ell+1}]\|_V \sum_{k=0}^\infty (1/s^\alpha)^k \leq \|\mathbb{E}[Y_\ell]\|_V / (s^\alpha - 1)$ for $\ell \geq 2$; cf. [134, p. 279]. Hence, $(s^\alpha - 1)\|\mathbb{E}[X_\ell] - \mathbb{E}[X]\|_V \leq \|\mathbb{E}[Y_\ell]\|_V \leq (1/s)^\alpha \|\mathbb{E}[Y_{\ell-1}]\|_V \leq (1/s)^{2\alpha} \|\mathbb{E}[Y_{\ell-2}]\|_V$ for $\ell \geq 3$. The above termination rule replaces the expectations of Y_ℓ, $Y_{\ell-1}$ and $Y_{\ell-2}$ with their sample means.
The differences $Y_\ell = X_\ell - X_{\ell-1}$ and the sum defining the MLMC estimator (see (4.3.1)) are computed using the coarse-to-fine operator [54, p. 159]. Algorithm 3 was implemented in `Python` and PDEs were solved using `FEniCS` [6, 220].

Model Problem

The following problem is an adaption of that considered by Barth, Schwab, and Zollinger [16, sect. 6.2]. We define $\mathcal{D} = (0, 1)$ and

$$\kappa(\omega)(x) = 5 + x + (2\sqrt{2}/\pi)\xi_1(\omega)\sin(\pi(x+1)/4), \quad b(\omega)(x) = 50 + 50\xi_2. \tag{4.5.8}$$

where ξ_1, $\xi_2 : \Omega \to \mathbb{R}$ are uniformly distributed over $[-1, 1]$ and dependent. For $x \in \bar{\mathcal{D}}$, we have

$$\mathbb{E}[y](x) = \sum_{k=0}^\infty \frac{(2\sqrt{2})^{2k}}{\pi^{2k}(2k+1)} \int_0^x \frac{c - 50z}{5+z} \Big(\frac{\sin(\pi(z+1)/4)}{5+z} \Big)^{2k} \mathrm{d}z, \tag{4.5.9}$$

where $c > 0$ is the solution to $\mathbb{E}[y](1) = 0$; cf. [16, p. 153].[5] Following [16, p. 153], we truncated the sum in (4.5.9) after five addends, and obtained $c \approx 24.25$ and $\mathbb{E}[y](1) \approx 4.0 \cdot 10^{-16}$ using `Mathematica`.

We verify the conditions of Lemmas 4.5.3 and 4.5.4. Fix $\omega \in \Omega$ and $x \in \bar{\mathcal{D}}$. We have $\kappa_{\min}(\omega)(x) \geq 5-2\sqrt{2}/\pi \approx 4.10$ and $\kappa_{\max}(\omega)(x) \leq 6+2\sqrt{2}/\pi \approx 6.90$. Since $\kappa(\omega)'(x) = 1+\cos(\pi(x+1)/4)/\sqrt{2}$, we find that $\|\kappa(\omega)'\|_{C(\bar{\mathcal{D}})} \leq 1 + 1/\sqrt{2}$ and $\|\kappa(\omega)\|_{C^1(\bar{\mathcal{D}})} \leq 7 + 2\sqrt{2}/\pi + 1/\sqrt{2} \approx 7.61$. Hence $\kappa \in L^\infty(\Omega; C^1(\bar{\mathcal{D}}))$. We also have $b \in L_\psi(\Omega; L^2(\mathcal{D}))$.

We define $V_\ell = Y_{h_\ell}$, following [54, sect. 0.4] and [16, sect. 6.2]. For $\ell \in \mathbb{N}$, we choose $h_\ell = 2^{-(\ell+1)}$ and divide $\bar{\mathcal{D}}$ into the intervals $[(k-1)h_\ell, kh_\ell]$ for $k = 1, \ldots, 1/h_\ell$. We define Y_{h_ℓ} as the space of continuous functions on $\bar{\mathcal{D}}$ that are zero at 0 and 1, and are affine polynomials on $[(k-1)2^{-\ell}, k2^{-\ell}]$ for $k = 1, \ldots, 1/h_\ell$.

From [54, Thm. 0.4.5] (with $\|\cdot\|_V = |\cdot|_{H^1(\mathcal{D})}$ [54, p. 5]) and its proof, we deduce $\inf_{v_h \in Y_{h_\ell}} |v - v_h|_{H^1(\mathcal{D})} \leq (1/\sqrt{2})h_\ell \|v''\|_{L^2(\mathcal{D})}$ for all $v \in H^2(\mathcal{D}) \cap H_0^1(\mathcal{D})$. Hence, Assumption (D5) is fulfilled, and the first two conditions in (4.3.2) hold with $\alpha = 1$ and $\beta = 2$.

The (theoretical) cost to generate a sample of $X_\ell = y_{h_\ell}$ is proportional to $1/h_\ell - 2$ since the solution of the discretized PDE (4.5.3) requires solving a tridiagonal linear system; see, e.g., [16, Rem. 4.6]. Consequently, the third condition in (4.3.2) holds with $\gamma = 1$.

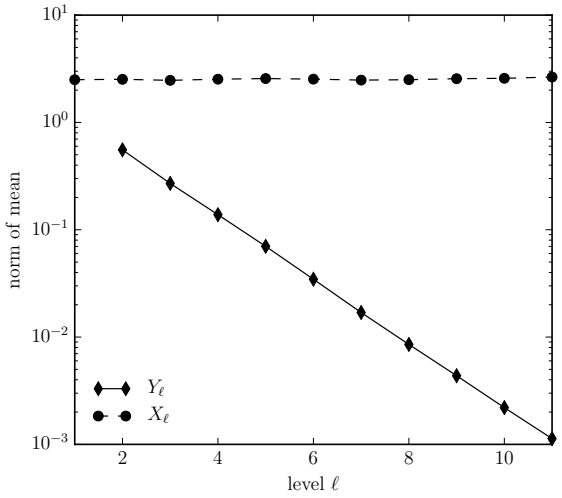
Figure 4.1 depicts several statistics for the problem (4.5.8). These statistics are adapted from those used by Giles [134, sect. 7]. To generate the results shown in Figure 4.1, we approximated expected values using $N = 500$ samples. Figure 4.1A depicts $\|\mathrm{E}^N[Y_\ell]\|_V$ and $\|\mathrm{E}^N[X_\ell]\|_V$, Figure 4.1B shows the estimates of the squared Luxemburg norms of $Y_\ell - \widetilde{Y}_\ell$ and $X_\ell - \widetilde{X}_\ell$, and Figure 4.1C depicts the average of the computation time per sample and level. Using least squares, we obtained the rates $\alpha \approx 0.99$, $\beta \approx 2.0$ and $\gamma \approx 0.51$, which empirically verify the conditions in (4.3.2), and Lemmas 4.5.2 and 4.5.4. The consistency error, visualized in Figure 4.1D, is the ratio of $(N^{1/2}/3)\|\mathrm{E}^N[Y_\ell] - \mathrm{E}^N[X_\ell] + \mathrm{E}^N[X_{\ell-1}]\|_V$ and the square root of $\mathrm{E}^N[\|Y_\ell - \widetilde{Y}_\ell\|_V^2] + \mathrm{E}^N[\|X_\ell - \widetilde{X}_\ell\|_V^2] + \mathrm{E}^N[\|X_{\ell-1} - \widetilde{X}_{\ell-1}\|_V^2]$; cf. [134, p. 22].

We compare `ProbMLMC` with the MC mean estimator described in section 4.3.2 using the same accuracy and reliability as those used for `ProbMLMC`. We refer to this MC estimator as `ProbMC`. Their computational cost is defined in sections 4.3.1 and 4.3.2. Here, L is the level computed by `ProbMLMC`. Figure 4.2 depicts the sample sizes and computational costs for several accuracies and reliabilities. Figure 4.2A shows that the cost of `ProbMC` increases significantly faster than that of `ProbMLMC` as ε decreases for the fixed reliability 0.95. For fixed accuracy, the normalized computational costs of `ProbMC` are a multiple of those of `ProbMLMC`; see Figure 4.2B. For both cases, `ProbMLMC` is computationally cheaper than `ProbMC`, and the results empirically verify the cost savings derived in section 4.3.2.

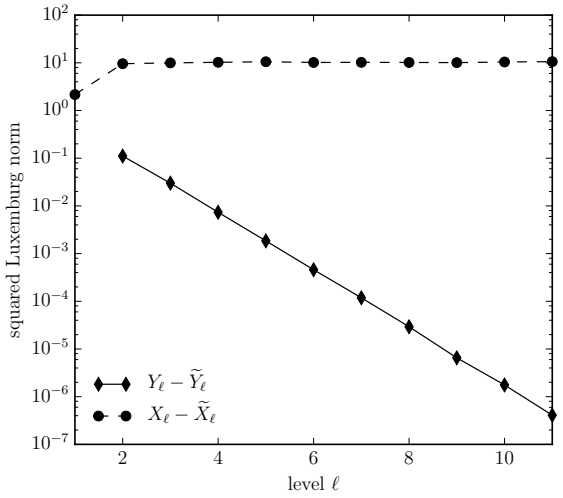
We empirically verify that `ProbMLMC` produces reliable mean estimates. We compare `ProbMLMC` with `StdMLMC`, the standard MLMC mean estimator analyzed by Bierig and Chernov [37, sect. 3], and with `VarMLMC`. `VarMLMC` is the same as `ProbMLMC` with the only difference that it uses estimates of the second moment of $Y_\ell - \mathbb{E}[Y_\ell]$ instead of the Luxemburg norm. The second moment of $Y_\ell - \mathbb{E}[Y_\ell]$ is estimated by that of $Y_\ell - \widetilde{Y}_\ell$. Lemma 4.2.10 ensures $\mathbb{E}[\|Y_\ell - \mathbb{E}[Y_\ell]\|_V^2] \leq \mathbb{E}[\|Y_\ell - \widetilde{Y}\|_V^2]$. Figure 4.3 depicts the relative frequency of $\|\mathrm{E}^{\mathrm{ML}}[X] - \mathbb{E}[X]\|_V > r$ as a function of $r \geq 0$ computed with 1000 independent simulations. It shows that `ProbMLMC` and `VarMLMC` yield a reliable mean estimate, while `StdMLMC` does not.

The number of samples used by Algorithm 3 is quite large, and the computational complexity of Algorithm 3 is unknown before executing the method. For simulations with (complex) PDEs, it may be more realistic that a fixed computational budget $C > 0$ is provided, and that we are interested in determining the number of levels L and samples, and the smallest constant $r > 0$

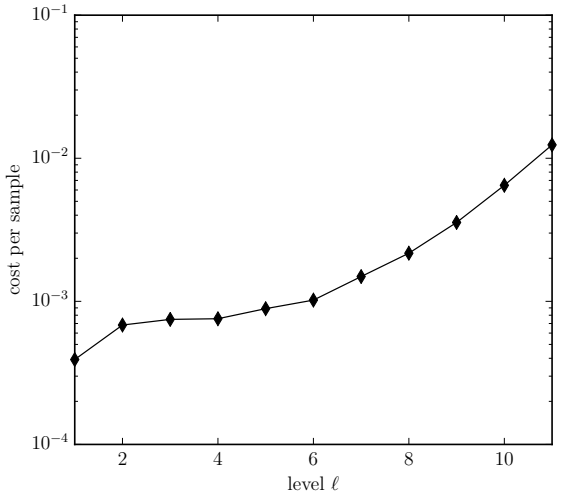
[5] The solution formula can be derived using the series expansion $1/(5 + x + z) = \sum_{k=1}^{\infty}(-z)^{k-1}/(x+5)^k$ for $z = (2\sqrt{2}/\pi)\xi_1(\omega)\sin(\pi(x+1)/4)$.



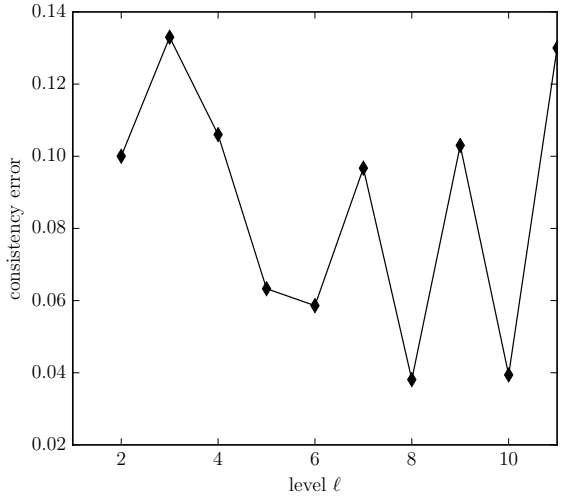
(A) $\|\mathrm{E}^N[Y_\ell]\|_V$ *and* $\|\mathrm{E}^N[X_\ell]\|_V$.



(B) *Estimates of the squared Luxemburg norm.*



(C) *Computational cost per sample.*



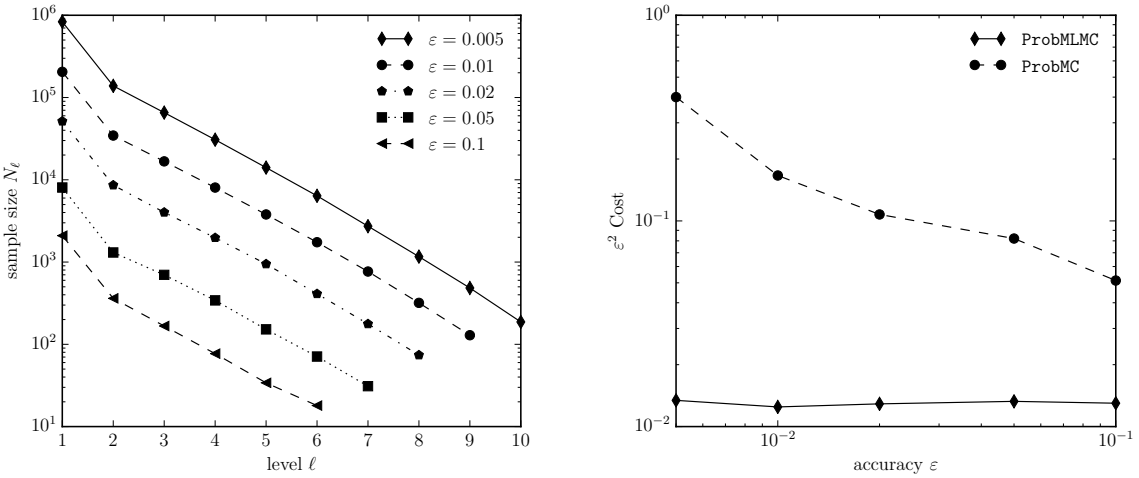
(D) *Consistency error.*

FIGURE 4.1: *Estimates of means, squared Luxemburg norm, computational costs and consistency error for* (4.5.8).

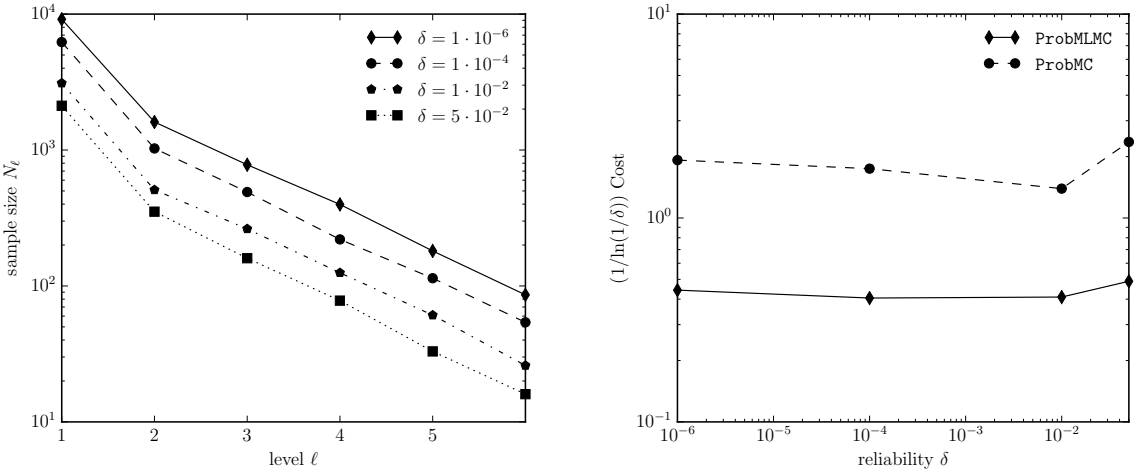
such that $\mathrm{Prob}(\|\mathrm{E}^{\mathrm{ML}}[X_L] - \mathbb{E}[X]\|_V \geq r) \leq \delta$, while ensuring that the overall simulation time does not exceed C. The tail bound (4.3.4) suggests obtaining the level L and the sample sizes N_ℓ as an optimal solution of

$$\min_{\substack{L\in\mathbb{N}\\ L\leq L_{\max}}} \|\mathbb{E}[X_L] - \mathbb{E}[X]\|_V + \min_{\substack{N_\ell\in\mathbb{N}\\ \ell=1,\dots,L}} \Big\{ (1+\sqrt{3\ln(1/\delta)})\Big(\sum_{\ell=1}^{L} \tfrac{\tau_\ell^2}{N_\ell}\Big)^{1/2} : \quad \sum_{\ell=1}^{L} N_\ell C_\ell \leq C \Big\}. \tag{4.5.10}$$

The second-stage program has an optimal solution if $C \geq C_1$. Let $r^* > 0$ be the optimal value and L^* be the optimal of (4.5.10). Since $V = H_0^1(\mathcal{D})$ is a Hilbert space, Theorem 4.2.15 ensures $\mathrm{Prob}(\|\mathrm{E}^{\mathrm{ML}}[X_{L^*}] - \mathbb{E}[X]\|_V \geq r^*) \leq \delta$. We approximated the second-stage program in (4.5.10) by relaxing each integer constraint to $(0,\infty)$. For fixed $L \in \mathbb{N}$, the optimal solution of the relaxation is $N_\ell = c(\tau_\ell^2/C_\ell)^{1/2}$ with $c = C/(\sum_{\ell=1}^{L}(\tau_\ell^2 C_\ell)^{1/2})$ if τ_ℓ, $C_\ell > 0$; cf. [134, p. 262]. The



(A) *Sample size N_ℓ as computed by* `ProbMLMC`*, and cost of* `ProbMC` *and* `ProbMLMC` *for $\delta = 0.05$, multiplied by ε^2.*



(B) *Sample size N_ℓ as computed by* `ProbMLMC`*, and cost of* `ProbMC` *and* `ProbMLMC` *for $\varepsilon = 0.1$, multiplied by $1/\ln(1/\delta)$.*

Figure 4.2: *Number of samples and computational costs* (4.5.8).

constants τ_ℓ^2, C_ℓ, and $\|\mathbb{E}[X_\ell] - \mathbb{E}[X]\|_V$ were estimated using the data depicted in Figure 4.1. We estimated $\|\mathbb{E}[X_1] - \mathbb{E}[X]\|_V$ using extrapolation. After rounding N_ℓ, the first-stage program in (4.5.10) can be solved approximately.
For our numerical experiments, we chose the computational budget C as a multiple of $\bar{C} = (1/L_{\max}) \sum_{\ell=1}^{L_{\max}} C_\ell$. Figure 4.4 depicts the sample sizes, and relative frequency of the deviation from the MLMC estimator to the true mean for 500 independent simulations. These results highlight the decrease of the sample sizes as the number of levels increases, and show that reliable mean estimates were obtained.

4.6 Conclusion and Discussion

We derived non-asymptotic, exponential bounds on the tail probabilities of the MLMC estimator applied to the mean estimation of certain Banach space-valued random variables in section 4.3. We required that the Banach spaces are either 2-uniformly smooth or that they are 2-uniformly

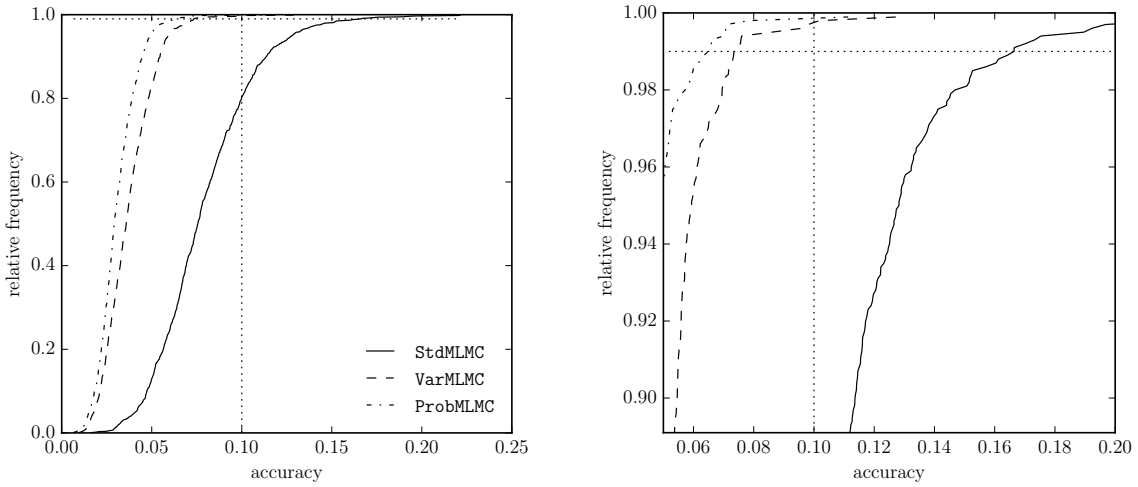


Figure 4.3: *Relative frequency of the deviation of the estimators to the true mean. The horizontal line is at* 0.98*, the desired reliability. The vertical line is at* 0.1*, the desired accuracy.*

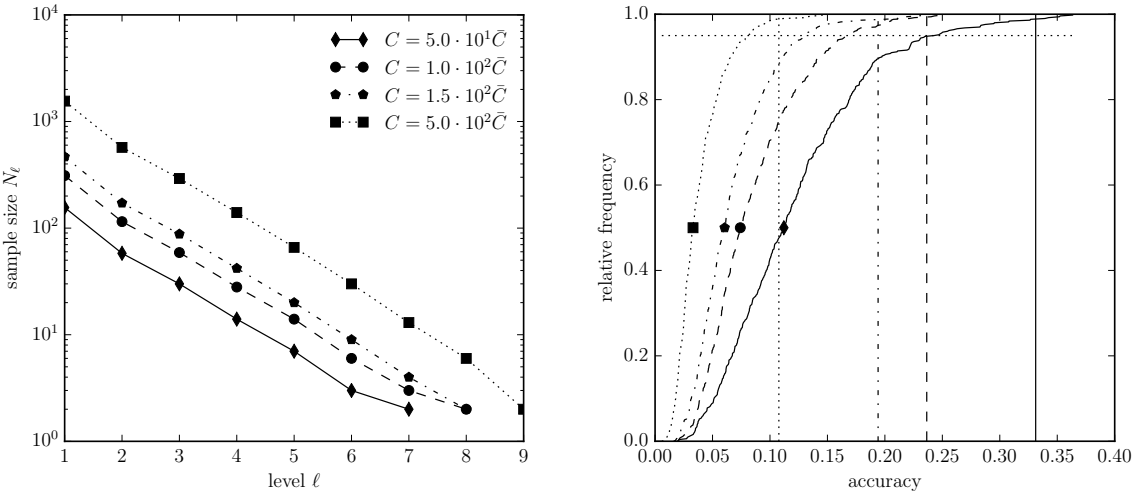


Figure 4.4: *Sample sizes computed using* (4.5.10)*, and relative frequency of the MLMC estimator's deviations to the true mean for* (4.5.8)*. The horizontal line is at* 0.95 *which is the desired reliability, and each vertical line is the (approximate) optimal value of* (4.5.10) *for the cost* C *which correspond to the desired accuracies. The legend applies to both subfigures.*

smooth after an equivalent renorming. For example, all Hilbert spaces and all Sobolev spaces consisting of at least square-integrable function are 2-uniformly smooth as demonstrated in section 4.2.2. Our analysis reveals that the number of samples on each level depend only logarithmically on the user-specified reliability if the multilevel corrections have sub-Gaussian tail behavior. The tail bounds established in section 4.2.4 depend on the geometry of the underlying Banach space through the smoothness constant κ. If κ is large, the tail bounds are informative to some extent. We established improved moment inequalities and tail bounds for certain Sobolev space-valued random sums in Proposition 4.2.14 and Theorem 4.2.18, respectively.

In section 4.5, we verified the assumptions used for our theory on linear elliptic PDEs with uniformly bounded diffusion coefficients and sub-Gaussian right-hand sides. Our framework is general enough to allow the application to stochastic obstacle problems under suitable assumptions on the random data. The sub-Gaussian tail behavior of the solutions to obstacle problems

and of the finite element approximation error can be established using the analysis developed by Bierig and Chernov [37] (see [37, eqns. (6.4) and (6.10)]).
It is possible to derive exponential tail bounds related to those in section 4.2.4 if the multilevel correction Y_ℓ satisfies, for some $\tau_\ell > 0$, $\mathbb{E}[\exp(\|Y_\ell - \mathbb{E}[Y_\ell]\|_V/\tau_\ell)] \leq \mathrm{e}$, which models sub-exponentiality and is weaker than the condition in (4.1.9). The corresponding Young's function is $\psi_1 : \mathbb{R} \to \mathbb{R}_+$, $\psi_1(x) = (\mathrm{e}^{|x|} - 1)/(\mathrm{e} - 1)$. The derivation of such tail bounds can be based on that by Juditsky and Nemirovski [166]. Giles [134, sect. 7.1] considers an elliptic PDE with a sub-exponential right-hand side.
We showed that the class of elliptic PDEs with log-normal random diffusion coefficients contains heavy-tailed solutions in section 4.5.2. We conjecture that most members of this class are heavy-tailed. Non-asymptotic tail bounds, weaker than exponential ones, may be established for these random variables using Nagaev–Fuk-type inequalities [356, Thm. 3.5.1], [117]. For Banach space-valued random variables, the Nagaev–Fuk inequality depends on unspecified constants which may be estimated for solutions to PDEs with log-normal random diffusion coefficients. In the case that the random variables X_ℓ can be approximated with light-tailed surrogates $\widehat{X}_\ell$ such that $\|\mathbb{E}[\widehat{X}_\ell] - \mathbb{E}[X]\|_V \approx \|\mathbb{E}[X_\ell] - \mathbb{E}[X]\|_V$, it would be possible to apply the framework developed in section 4.3.1. The Gaussian random variables entering the log-normal diffusion coefficient may be approximated using truncation; see, e.g., [323, sect. A661] and [245, p. 982] for approximations of log-normal random variables. An open problem is whether these approximations result in $\|\mathbb{E}[\widehat{X}_\ell] - \mathbb{E}[X]\|_V \approx \|\mathbb{E}[X_\ell] - \mathbb{E}[X]\|_V$.
The mean-squared analysis for MLMC mean estimators, applied to solutions of linear elliptic PDEs with log-normal diffusion coefficients, exploits the fact that centered second moments of the multilevel corrections decrease sufficiently fast to zero [69, 70, 311]. However, under mild assumptions, higher-order moments also decrease with increasing levels [70, Thm. 3.9], [311, Thm. 2.2]—a property that may be exploited for the development of robust MLMC mean estimators.
MLMC mean estimators with sub-Gaussian behavior for heavy-tailed random variables can be built on computing the geometric median of a moderate number of independent MLMC mean estimators. Minsker [236] shows that the geometric median of independent estimators has significantly greater reliability than each individual estimator (see also Nemirovski and Yudin [246, p. 244]). Another approach could be to replace the sample means on each level with a robust mean estimator, that is, by a mean estimator with (nearly) sub-Gaussian performance. The development of such estimators is an active research field for univariate and multivariate random variables; see, e.g., [67, 222, 236].

4.7 Proofs and Supplementary Materials

4.7.1 Uniform Smoothness of Sobolev Spaces

We prove the fact that $W^{s,p}(\mathcal{D})$ is $(2, p-1)$-smooth if $s \in \mathbb{N}_0$, $2 \leq p < \infty$ and $\mathcal{D} \subset \mathbb{R}^d$ is a bounded domain. The proof uses Hanner's inequality [145, Thm. 1], [13, Thm. 2]. For each fixed $2 \leq p < \infty$, Hanner's inequality ensures, for every f, $g \in L^p(\mathcal{D})$,

$$\|f+g\|_{L^p(\mathcal{D})}^p + \|f-g\|_{L^p(\mathcal{D})}^p \leq (\|f\|_{L^p(\mathcal{D})} + \|g\|_{L^p(\mathcal{D})})^p + (\|f\|_{L^p(\mathcal{D})} - \|g\|_{L^p(\mathcal{D})})^p. \tag{4.7.1}$$

Proof of Proposition 4.2.7. We verify (4.2.2). Fix $s \in \mathbb{N}_0$, $2 \leq p < \infty$, and x, $y \in W^{s,p}(\mathcal{D})$. We define $(v_\alpha)_{|\alpha| \leq s}$ and $(w_\alpha)_{|\alpha| \leq s}$ by $v_\alpha = \|\mathrm{D}^\alpha x\|_{L^p(\mathcal{D})}$ and $w_\alpha = \|\mathrm{D}^\alpha y\|_{L^p(\mathcal{D})}$ for $|\alpha| \leq s$, respectively. Here, D^α is the weak derivative of order α, and $\alpha \in \mathbb{N}_0^d$ is a multiindex (see pp. vii–viii). Using the definition of the Sobolev norm $\|\cdot\|_{W^{s,p}}$ (see p. viii), we have $\|v\|_p = \|x\|_{W^{s,p}(\mathcal{D})}$

and $\|w\|_p = \|y\|_{W^{s,p}(\mathcal{D})}$. For each $\alpha \in \mathbb{N}_0^d$ with $|\alpha| \le s$, we have $\mathrm{D}^\alpha x$, $\mathrm{D}^\alpha y \in L^p(\mathcal{D})$ and, hence, Hanner's inequality (4.7.1), applied with $f = \mathrm{D}^\alpha x$ and $g = \mathrm{D}^\alpha y$, ensures

$$\begin{aligned}\|x+y\|_{W^{s,p}(\mathcal{D})}^p + \|x-y\|_{W^{s,p}(\mathcal{D})}^p &= \sum_{|\alpha|\le s} \|\mathrm{D}^\alpha[x+y]\|_{L^p(\mathcal{D})}^p + \|\mathrm{D}^\alpha[x-y]\|_{L^p(\mathcal{D})}^p \\ &\le \sum_{|\alpha|\le s} (v_\alpha + w_\alpha)^p + |v_\alpha - w_\alpha|^p \\ &= \|v+w\|_p^p + \|v-w\|_p^p.\end{aligned}$$

Using [13, Thm. 1] (see also [262, eq. (10.37)]), we obtain $(\|v+w\|_p^p/2 + \|v-w\|_p^p/2)^{2/p} \le \|v\|_p^2 + (p-1)\|w\|_p^2$. Putting together the pieces, we find that

$$\begin{aligned}(\|x+y\|_{W^{s,p}(\mathcal{D})}^p/2 + \|x-y\|_{W^{s,p}(\mathcal{D})}^p/2)^{2/p} &\le (\|v+w\|_p^p/2 + \|v-w\|_p^p/2)^{2/p} \\ &\le \|v\|_p^2 + (p-1)\|w\|_p^2 \\ &= \|x\|_{W^{s,p}(\mathcal{D})}^2 + (p-1)\|y\|_{W^{s,p}(\mathcal{D})}^2.\end{aligned}$$

Using $2 \le p < \infty$ and $(a^2/2 + b^2/2)^{1/2} = 2^{-1/2}\|(a,b)\|_2 \le 2^{-1/2}2^{1/2-1/p}\|(a,b)\|_p = (a^p/2 + b^p/2)^{1/p}$ for $a = \|x+y\|_{W^{s,p}(\mathcal{D})}$ and $b = \|x-y\|_{W^{s,p}(\mathcal{D})}$, we conclude that $W^{s,p}(\mathcal{D})$ is $(2, p-1)$-smooth. □

4.7.2 Renorming

We show that the renoming lemma by Juditsky and Nemirovski [166, Lem. 3] remains valid for infinite-dimensional Banach spaces. A standard reference on renorming is the monograph [95].

Lemma 4.7.1. *If $(V, \|\cdot\|_V)$ is $(2,\kappa)$-quasi-smooth Banach space, then there exists a norm $|||\cdot|||_V$ on V such that $|||\cdot|||_V^2$ is $(2,\kappa)$-smooth and*

$$\|x\|_V^2 \le |||x|||_V^2 \le 2\|x\|_V^2 \quad \text{for all} \quad x \in V. \tag{4.7.2}$$

Our proof of Lemma 4.7.1, which is inspired by that of [166, Lem. 3], requires some facts from functional analysis. Throughout the following sections, $\|\cdot\|_{V^*}$ is the dual norm to $\|\cdot\|_V$, where $(V, \|\cdot\|_V)$ is a Banach space.

Lemma 4.7.2 ([155, Thm. on p. 155])**.** *If $(V, \|\cdot\|_V)$ is a Banach space, then each norm on V^* that is equivalent to $\|\cdot\|_{V^*}$ is a dual norm if and only if V is reflexive.*

Lemma 4.7.3 ([155, p. 154])**.** *Let V be a Banach space, and let $\|\cdot\|_V$ and $|||\cdot|||_V$ be norms on V such that, for some α, $\beta > 0$, we have $\alpha\|x\|_V \le |||x|||_V \le \beta\|x\|_V$ for all $x \in V$. Then*

$$(1/\beta)\|f\|_{V^*} \le |||f|||_{V^*} \le (1/\alpha)\|f\|_{V^*} \quad \text{for all} \quad f \in V^*.$$

Lemma 4.7.4. *If $(V, \|\cdot\|_V)$ is a Banach space and $\kappa \ge 1$, then $\|\cdot\|_V^2$ is $(2,\kappa)$-smooth if and only if*

$$\|f+g\|_{V^*}^2 + \|f-g\|_{V^*}^2 \ge 2\|f\|_{V^*}^2 + 2(1/\kappa)\|g\|_{V^*}^2 \quad \text{for all} \quad f, g \in V^*.$$

Proof. The assertions follow from an application of [13, Lem. 5]. □

Proof of Lemma 4.7.1. The proof is inspired by that of [166, Lem. 3]. The main idea is to construct a dual norm on V^* that fulfills the inequality in Lemma 4.7.4. Since $(V, \|\cdot\|_V)$ is

$(2,\kappa)$-quasi-smooth, there exists a norm $|\cdot|_V$ on V, and $\bar{\kappa} \in [1,\kappa]$ such that $|\cdot|_V^2$ is $(2,\bar{\kappa})$-smooth and

$$\|x\|_V^2 \le |x|_V^2 \le \mu \|x\|_V^2 \quad \text{for all} \quad x \in V \quad \text{with} \quad \mu = \kappa/\bar{\kappa}. \tag{4.7.3}$$

If $\mu \in [1,2]$, we choose $|||\cdot|||_V^2 = |\cdot|_V^2$, which is $(2,\kappa)$-smooth since $\bar{\kappa} \le \kappa$. For the remainder of the proof, let $\mu > 2$ be arbitrary. Lemma 4.7.3 yields

$$(1/\mu)\|g\|_{V^*}^2 \le |g|_{V^*}^2 \le \|g\|_{V^*}^2 \quad \text{for all} \quad g \in V^*. \tag{4.7.4}$$

We define the norm $|||\cdot|||_{V^*}$ on V^* by $|||\cdot|||_{V^*}^2 = \gamma|\cdot|_{V^*}^2 + (1-\gamma)\|\cdot\|_{V^*}^2$, where $\gamma = \mu/(2(\mu-1))$ with $1/2 \le \gamma < 1$. Indeed, $\mathrm{g} : [2,\infty) \to \mathbb{R}$ defined by $\mathrm{g}(z) = \frac{1}{2(1-1/z)}$ is differentiable on $(2,\infty)$ with $\mathrm{g}'(z) = -\frac{1}{2(z-1)^2} < 0$ for all $z \in (2,\infty)$. Since $\gamma = \mathrm{g}(\mu)$, $\mu > 2$, $\mathrm{g}(2) = 1$ and $\mathrm{g}(z) \to 1/2$ as $z \to \infty$, we have $1/2 \le \gamma < 1$.
Using (4.7.4) and $\gamma/\mu + 1 - \gamma = 1/2$, we obtain for all $g \in V^*$,

$$(1/2)\|g\|_{V^*}^2 = (\gamma/\mu + 1 - \gamma)\|g\|_{V^*}^2 \le |||g|||_{V^*}^2 = \gamma|g|_{V^*}^2 + (1-\gamma)\|g\|_{V^*}^2 \le \|g\|_{V^*}^2. \tag{4.7.5}$$

Since V is reflexive [50, Thm. 5.1.20], Lemma 4.7.2 implies that $|||\cdot|||_{V^*}$ is a dual norm, that is, $|||\cdot|||_{V^*}$ is the dual norm of some norm $|||\cdot|||_V$ on V. Reflexivity of V also implies that the dual norms of $|||\cdot|||_{V^*}$ and of $\|\cdot\|_{V^*}$ are norms on V. Hence, Lemma 4.7.3 and (4.7.5) yield (4.7.2).
It must yet be shown that $|||\cdot|||_V^2$ is $(2,\kappa)$-smooth. Fix f, $g \in V^*$. The convexity and continuity of $\|\cdot\|_{V^*}^2$ and the fact that V^* is a Banach space [155, p. 120] ensure $\|f+h\|_{V^*}^2 \ge \|f\|_{V^*}^2 + \langle y, h\rangle_{(V^*)^*,V^*}$ for some subgradient $y \in (V^*)^*$ of $\|\cdot\|_{V^*}^2$ at f, and every $h \in V^*$; see, e.g., [46, Prop. 2.126 (v)]. Hence, $\|f+g\|_{V^*}^2 + \|f-g\|_{V^*}^2 \ge 2\|f\|_{V^*}^2$. Combined with Lemma 4.7.4 applied to $|\cdot|_{V^*}$ and the definition of $|||\cdot|||_{V^*}$, we find that

$$\begin{aligned} |||f+g|||_{V^*}^2 + |||f-g|||_{V^*}^2 &= \gamma\big(|f+g|_{V^*}^2 + |f-g|_{V^*}^2\big) + (1-\gamma)\big(\|f+g\|_{V^*}^2 + \|f-g\|_{V^*}^2\big) \\ &\ge 2\gamma|f|_{V^*}^2 + 2(\gamma/\bar{\kappa})|g|_{V^*}^2 + 2(1-\gamma)\|f\|_{V^*}^2 \\ &= 2|||f|||_{V^*}^2 + 2(\gamma/\bar{\kappa})|g|_{V^*}^2. \end{aligned} \tag{4.7.6}$$

Using (4.7.4), we obtain

$$|||g|||_{V^*}^2 = \gamma|g|_{V^*}^2 + (1-\gamma)\|g\|_{V^*}^2 \le \big(\gamma + (1-\gamma)\mu\big)|g|_{V^*}^2. \tag{4.7.7}$$

Since $\gamma = \mu/(2(\mu-1))$ and $\mu > 2$, we have $\mu\gamma/(\gamma + (1-\gamma)\mu) = \mu/(\mu-1) \ge 1$. Combined with (4.7.6), (4.7.7), and $\bar{\kappa} = \kappa/\mu$ (see (4.7.3)), we conclude that

$$\begin{aligned} |||f+g|||_{V^*}^2 + |||f-g|||_{V^*}^2 &\ge 2|||f|||_{V^*}^2 + \tfrac{2\mu\gamma}{\kappa}|g|_{V^*}^2. \\ &= 2|||f|||_{V^*}^2 + \tfrac{2}{\kappa}\tfrac{\mu\gamma}{\gamma+(1-\gamma)\mu}|||g|||_{V^*}^2 \\ &\ge 2|||f|||_{V^*}^2 + \tfrac{2}{\kappa}|||g|||_{V^*}^2. \end{aligned}$$

Hence, Lemma 4.7.4 implies that $|||\cdot|||_V^2$ is $(2,\kappa)$-smooth. □

4.7.3 Proofs of Bounds on the Second Moment

We prove Theorem 4.2.11 using Lemmas 4.7.5 and 4.7.6.

Lemma 4.7.5. *If $(V, \|\cdot\|_V)$ is $(2,\kappa)$-smooth, then $g = \|\cdot\|_V^2$ is uniformly Fréchet differentiable, and $|\mathrm{D}g(y)[v]| \le 2\|y\|_V\|v\|_V$ for all y, $v \in V$.*

Proof. Owing to (4.2.2) and the convexity of $g = \|\cdot\|_V^2$, the function g is uniformly Fréchet differentiable [50, Prop. 4.2.14]. From [14, Ex. 2.32], we deduce $|\mathrm{D}g(y)[v]| \leq \|\mathrm{D}g(y)\|_{V^*}\|v\|_V \leq 2\|y\|_V\|v\|_V$ for all y, $v \in V$. □

We state an essentially known characterization of $(2,\kappa)$-smoothness.

Lemma 4.7.6. *Let $(V, \|\cdot\|_V)$ be a Banach space and $\kappa \geq 1$. Then $\|\cdot\|_V^2$ is $(2,\kappa)$-smooth if and only if $g = \|\cdot\|_V^2$ is Fréchet differentiable, and*

$$g(x+y) \leq g(x) + \mathrm{D}g(x)[y] + \kappa g(y) \quad \textit{for all} \quad x,\, y \in V. \tag{4.7.8}$$

Proof. If $\|\cdot\|_V^2$ is $(2,\kappa)$-smooth, then [50, Prop. 4.2.14] ensures the (uniform) Fréchet differentiability of $\|\cdot\|_V^2$, and the proof of [164, Lem. 2.2] implies (4.7.8). To establish the converse, we fix x, $y \in V$. We choose once x, y in (4.7.8) and once x, $-y$ in (4.7.8), and add the two inequalities. Combined with the linearity of $\mathrm{D}g(x)$, we obtain (4.2.2). Hence, $(V, \|\cdot\|_V)$ is $(2,\kappa)$-smooth. □

Proof of Theorem 4.2.11. The proof is inspired by that of [256, Prop. 2.5], and the statements presented in [212, Rem. 2.3] and [166, p. 4].
Let $(\xi_j)_{j\in\mathbb{N}_0}$ be adapted to the filtration $(\mathcal{F}_j)_{j\in\mathbb{N}_0} \subset \mathcal{F}$. We define $S_0 = 0$, $S_j = \xi_1 + \cdots + \xi_j$ for each $j \in \mathbb{N}$, and $g = \|\cdot\|_V^2$. The martingale $(S_j)_{j\in\mathbb{N}_0}$ is adapted to $(\mathcal{F}_j)_{j\in\mathbb{N}_0}$ [159, Ex. 3.1.7]. Owing to Lemma 4.7.6, we have for each $j = 1, 2, \ldots$, w.p. 1,

$$\|S_j + x\|_V^2 \leq \|S_{j-1} + x\|_V^2 + \mathrm{D}g(S_{j-1} + x)[\xi_j] + \kappa\|\xi_j\|_V^2. \tag{4.7.9}$$

Fix $j \in \mathbb{N}$. Using $\mathbb{E}[\xi_j \mid \mathcal{F}_{j-1}] = 0$, $\xi_j \in L^2(\Omega; V)$, $|\mathrm{D}g(S_{j-1} + x)[\xi_j]| \leq 2\|S_{j-1} + x\|_V\|\xi_j\|_V$ (see Lemma 4.7.5), the tower property [159, Prop. 2.6.33], and [159, Prop. 2.6.31], we obtain

$$\mathbb{E}[\mathrm{D}g(S_{j-1} + x)[\xi_j]] = \mathbb{E}[\mathbb{E}[\mathrm{D}g(S_{j-1} + x)[\xi_j] \mid \mathcal{F}_{j-1}]] = \mathbb{E}[\mathrm{D}g(S_{j-1} + x)\mathbb{E}[\xi_j \mid \mathcal{F}_{j-1}]] = 0.$$

Taking expectations in (4.7.9), we find that

$$\mathbb{E}[\|S_j + x\|_V^2] - \mathbb{E}[\|S_{j-1} + x\|_V^2] \leq \kappa\mathbb{E}[\|\xi_j\|_V^2] \quad \text{for} \quad j = 1, 2, \ldots.$$

Combined with the telescoping sum and $\mathbb{E}[\|S_0 + x\|_V^2] = \|x\|_V^2$, we obtain (4.2.4). □

Proof of Corollary 4.2.12. By assumption, there exists a norm $|||\cdot|||_V$ on V such that $|||\cdot|||_V^2$ is $(2,\bar\kappa)$-smooth with $\bar\kappa \in [1,\kappa]$ and $\|\cdot\|_V^2 \leq |||\cdot|||_V^2 \leq (\kappa/\bar\kappa)\|\cdot\|_V^2$. Theorem 4.2.11 yields $\mathbb{E}[|||\xi_1 + \cdots + \xi_N|||_V^2] \leq \bar\kappa\sum_{i=1}^N \mathbb{E}[|||\xi_i|||_V^2]$. Putting together the pieces, we obtain (4.2.5). □

Proof of Corollary 4.2.13. We define $(S_j)_{j\in\mathbb{N}_0}$ with $S_0 = 0$ and such that S_j is the sum of the first j addends of $\sum_{\ell=1}^L\sum_{i=1}^{N_\ell} \xi_{\ell,i}/N_\ell$, and $S_{n+1} = S_N$ for all $n \geq N$, where $N = \sum_{\ell=1}^L N_\ell$. The independence of $\xi_{\ell,i} \in L^2(\Omega; V)$ $(i = 1, \ldots, N_\ell,\, \ell = 1, \ldots, L)$ implies that $(S_j)_{j\in\mathbb{N}_0}$ is a (stopped) martingale adapted to the natural filtration [159, Ex. 3.1.4]. Hence, $(d_j)_{j\in\mathbb{N}_0}$ with $d_0 = 0$ and $d_j = S_j - S_{j-1}$ for $j \in \mathbb{N}$ is a martingale-difference. Applying Theorem 4.2.11 to $(d_j)_{j\in\mathbb{N}_0}$ yields the bound (4.2.7). An application of Corollary 4.2.12 gives (4.2.6). □

We make use of Lemma 4.7.7 to prove Proposition 4.2.14.

Lemma 4.7.7. *If $\xi_k : \Omega \to \mathbb{R}$ are independent and sub-Gaussian with parameter $\tau_k > 0$, and $a_k \in \mathbb{R}$ for $k = 1, \ldots, K$, then for each $r > 0$, $\mathbb{E}[|\sum_{k=1}^K a_k\xi_k|^r] \leq 2(r/\mathrm{e})^{r/2}(\sum_{k=1}^K a_k^2\tau_k^2)^{r/2}$. If, in addition, ξ_k are Gaussian with variance τ_k^2, then for each $r > 0$, $\mathbb{E}[|\sum_{k=1}^K a_k\xi_k|^r] = \mathbb{E}[|\xi_1/\tau_1|^r](\sum_{k=1}^K a_k^2\tau^2)^{r/2}$.*

Proof. The first claim follows from applications of [57, Thm. 1.2, Lems. 1.4 and 1.7 (sect. 1.1)]. We prove the second assertion. Define $Z = \sum_{k=1}^{K} a_k \xi_k$. Since ξ_k are independent, mean-zero Gaussian random variables with variance τ_k^2, we have $\mathbb{E}[Z^2] = \sum_{k=1}^{K} a_k^2 \tau_k^2$. Hence, $Z/\mathbb{E}[Z^2]^{1/2}$ is a mean-zero Gaussian random variable with unit variance. Consequently, $\mathbb{E}[|Z|^r] = \mathbb{E}[|Z/\mathbb{E}[Z^2]^{1/2}|^r]\mathbb{E}[Z^2]^{r/2} = \mathbb{E}[|\xi_1/\tau_1|^r](\sum_{k=1}^{K} a_k^2\tau_k^2)^{r/2}$. □

Proof of Proposition 4.2.14. We define $Z = \sum_{k=1}^{K} \xi_k \phi_k$. Using the definition of $\|\cdot\|_{W^{s,p}(\mathcal{D})}$ (see p. viii), we have $\mathbb{E}[\|Z\|_{W^{s,p}(\mathcal{D})}^p] = \mathbb{E}[\sum_{|\alpha|\leq s} \|\mathrm{D}^\alpha Z\|_{L^p(\mathcal{D})}^p]$. Since $s \in \mathbb{N}_0$ and $t \mapsto |t|^p$ is nonnegative, Fubini's theorem ensures $\mathbb{E}[\|Z\|_{W^{s,p}(\mathcal{D})}^p] = \sum_{|\alpha|\leq s} \int_{\mathcal{D}} \mathbb{E}[|\mathrm{D}^\alpha Z(x)|^p]\mathrm{d}x$.

Since $\phi_k \in W^{s,p}(\mathcal{D})$ for $k = 1, \ldots, K$, there exists a set $\widehat{\mathcal{D}} \subset \mathcal{D}$ of full measure such that $\mathrm{D}^\alpha \phi_k(x) \in \mathbb{R}$ for $k = 1, \ldots, K$, $|\alpha| \leq s$, and for all $x \in \widehat{\mathcal{D}}$. Now, we fix $x \in \widehat{\mathcal{D}}$ and $s \in \mathbb{N}_0$. Using Lemma 4.7.7 and the linearity of D^α [1, p. 21], we find that

$$\mathbb{E}\Big[\Big|\mathrm{D}^\alpha\Big(\sum_{k=1}^{K} \xi_k \phi_k(x)\Big)\Big|^p\Big] = \mathbb{E}\Big[\Big|\sum_{k=1}^{K} \xi_k \mathrm{D}^\alpha \phi_k(x)\Big|^p\Big] \leq 2(p/\mathrm{e})^{p/2}\Big(\sum_{k=1}^{K} \tau_k^2 (\mathrm{D}^\alpha \phi_k(x))^2\Big)^{p/2}.$$

Putting together the pieces and using the definition of T_K provided in (4.2.8), we conclude that

$$\mathbb{E}[\|Z\|_{W^{s,p}(\mathcal{D})}^p] \leq 2(p/\mathrm{e})^{p/2} \sum_{|\alpha|\leq s} \int_{\mathcal{D}} \Big[\sum_{k=1}^{K} \tau_k^2 (\mathrm{D}^\alpha \phi_k(x))^2\Big]^{p/2} \mathrm{d}x = 2(p/\mathrm{e})^{p/2} T_K^p.$$

If, in addition, ξ_k are Gaussian, then the above computations combined with Lemma 4.7.7 yield $\mathbb{E}[\|Z\|_{W^{s,p}(\mathcal{D})}^p] = \mathbb{E}[|\xi_1/\tau_1|^p] T_K^p$. □

4.7.4 Proofs of Exponential Tail Bounds

We prove Theorems 4.2.15, 4.2.16 and 4.2.18 and Corollary 4.2.17. Theorems 4.7.8 and 4.7.9 are used to establish Theorem 4.2.15.

Theorem 4.7.8 ([259, Thm. 2], [356, Thm. 3.3.3])**.** *If $(V, \|\cdot\|_V)$ is a separable Banach space, $\xi_j \in L^1(\Omega; V)$ for $j = 1, \ldots, N \in \mathbb{N}$ are mean-zero and independent, and $S_N = \xi_1 + \cdots + \xi_N$, then for all $\lambda \geq 0$, $\mathbb{E}[\exp(\lambda \|S_N\|_V)] \leq \exp(\lambda \mathbb{E}[\|S_N\|_V]) \prod_{j=1}^{N} \mathbb{E}[\exp(\lambda \|\xi_j\|_V) - \lambda \|\xi_j\|_V]$.*

Theorem 4.7.9 ([258, Thm. 1.2])**.** *If the hypotheses of Theorem 4.7.8 hold and $\|\xi_j\|_{L^\infty(\Omega;V)} \leq \tau_j$ for $j = 1, \ldots, N \in \mathbb{N}$, then for all $r > 0$, $\mathrm{Prob}(\|S_N\|_V \geq \mathbb{E}[\|S_N\|_V] + r) \leq \exp(-r^2/(2T_N^2))$, where $T_N = (\sum_{j=1}^{N} \tau_j^2)^{1/2}$.*

We restate Lemmas 3.6.5 and 3.6.6.

Lemma 3.6.5. *If $\xi \in L^0(\Omega; V)$ and $\mathbb{E}[\exp(\sigma^{-2}\|\xi\|_V^2)] \leq \mathrm{e}$ for some $\sigma > 0$, then*

$$\mathbb{E}[\exp(\lambda\|\xi\|_V) - \lambda\|\xi\|_V] \leq \exp(3\lambda^2\sigma^2/4) \quad \text{for all} \quad \lambda \geq 0. \tag{4.7.10}$$

Lemma 3.6.6. *If a, $b > 0$, then $\min_{\lambda>0} -a\lambda + b\lambda^2 = -a^2/(4b)$.*

Proof of Theorem 4.2.15. Fix ε, $\lambda > 0$, $L \in \mathbb{N}$, $i \in \mathbb{N}$, $N_\ell \in \mathbb{N}$, for $\ell = 1, \ldots, L$, and $\ell \in \{1, \ldots, L\}$. We define $(S_j)_{j\in\mathbb{N}_0}$ as in the proof of Corollary 4.2.13, that is, $S_j = 0$ and S_j is the sum of the first j addends of $\sum_{\ell=1}^{L}\sum_{i=1}^{N_\ell} \xi_{\ell,i}/N_\ell$, and $N = \sum_{\ell=1}^{L} N_\ell$. We establish the assertions through applying Theorems 4.7.8 and 4.7.9 to S_N combined with a Chernoff-type approach. Jensen's inequality, Corollary 4.2.13 and Lemma 4.2.1 imply

$$(\mathbb{E}[\|S_N\|_V])^2 \leq \mathbb{E}[\|S_N\|_V^2] \leq \kappa \sum_{\ell=1}^{L}\sum_{i=1}^{N_\ell} \frac{\mathbb{E}[\|\xi_{\ell,i}\|_V^2]}{N_\ell^2} \leq \kappa \sum_{\ell=1}^{L} \frac{\tau_\ell^2}{N_\ell}. \tag{4.7.11}$$

Since $\mathbb{E}[\exp(\tau_\ell^{-2}\|\xi_{\ell,i}\|_V^2)] \leq \mathrm{e}$, Lemma 3.6.5 ensures $\mathbb{E}[\mathrm{e}^{\lambda\|\xi_{\ell,i}/N_\ell\|_V} - \lambda\|\xi_{\ell,i}/N_\ell\|_V] \leq \mathrm{e}^{3\lambda^2\tau_\ell^2/(4N_\ell^2)}$. Combined with (4.7.11), Markov's inequality and Theorem 4.7.8, we find that

$$\begin{aligned}
\mathrm{Prob}\Big(\|S_N\|_V \geq \Big(\kappa\sum_{\ell=1}^{L}\tfrac{\tau_\ell^2}{N_\ell}\Big)^{1/2} + \varepsilon\Big) &\leq \mathrm{Prob}(\|S_N\|_V \geq \mathbb{E}[\|S_N\|_V] + \varepsilon) \\
&\leq \mathrm{e}^{-\lambda\varepsilon}\mathbb{E}[\mathrm{e}^{\lambda\|S_N\|_V - \lambda\mathbb{E}[\|S_N\|_V]}] \\
&\leq \mathrm{e}^{-\lambda\varepsilon}\prod_{j=1}^{N}\mathbb{E}[\mathrm{e}^{\lambda\|S_j - S_{j-1}\|_V} - \lambda\|S_j - S_{j-1}\|_V] \\
&= \mathrm{e}^{-\lambda\varepsilon}\prod_{\ell=1}^{L}\prod_{i=1}^{N_\ell}\mathbb{E}[\mathrm{e}^{\lambda\|\xi_{\ell,i}/N_\ell\|_V} - \lambda\|\xi_{\ell,i}/N_\ell\|_V] \\
&\leq \mathrm{e}^{-\lambda\varepsilon + \sum_{\ell=1}^{L}\sum_{i=1}^{N_\ell} 3\lambda^2\tau_\ell^2/(4N_\ell^2)} \\
&= \mathrm{e}^{-\lambda\varepsilon + \sum_{\ell=1}^{L} 3\lambda^2\tau_\ell^2/(4N_\ell)}.
\end{aligned}$$

Minimizing this bound over $\lambda > 0$, applying Lemma 3.6.6, and choosing $\varepsilon^2 = r^2\sum_{\ell=1}^{L}(\tau_\ell^2/N_\ell)$ yields (4.2.9).
If, in addition, $\|\xi_{\ell,i}\|_{L^\infty(\Omega;V)} \leq \tau_\ell$ for i, $\ell = 1, 2, \ldots$, then

$$\sum_{j=1}^{N}\|S_j - S_{j-1}\|_{L^\infty(\Omega;V)}^2 = \sum_{\ell=1}^{L}\sum_{i=1}^{N_\ell}\|\xi_{\ell,i}/N_\ell\|_{L^\infty(\Omega;V)}^2 \leq \sum_{\ell=1}^{L}\sum_{i=1}^{N_\ell}\tau_\ell^2/N_\ell^2 = \sum_{\ell=1}^{L}\tau_\ell^2/N_\ell. \tag{4.7.12}$$

Combined with the above computations and an application of Theorem 4.7.9 instead of Theorem 4.7.8, we deduce the second assertion. □

Proof of Corollary 4.2.17. The proof is based on that of [166, Thm. 2.1]. Lemma 4.7.1 implies the existence of a norm $|||\cdot|||_V$ on V such that $|||\cdot|||_V^2$ is $(2,\kappa)$-smooth and $\|\cdot\|_V^2 \leq |||\cdot|||_V^2 \leq 2\|\cdot\|_V^2$ (see (4.7.2)). Combined with $\mathbb{E}[\exp(\tau_\ell^{-2}\|\xi_{\ell,i}\|_V^2)] \leq \mathrm{e}$, we get $\mathbb{E}[\exp((\sqrt{2}\tau_\ell)^{-2}|||\xi_{\ell,i}|||_V^2)] \leq \mathrm{e}$. Now, Theorem 4.2.15 yields $\mathrm{Prob}(\|\sum_{\ell=1}^{L}\frac{1}{N_\ell}\sum_{i=1}^{N_\ell}\xi_{\ell,i}\|_V \geq (\sqrt{\kappa}+r)(\sum_{\ell=1}^{L}2\tau_\ell^2/N_\ell)^{1/2}) \leq \exp(-r^2/3)$ (see (4.2.9)). Since $\|\cdot\|_V \leq |||\cdot|||_V$, we have $\mathrm{Prob}(\|S_N\|_V \geq \varepsilon) \leq \mathrm{Prob}(|||S_N|||_V \geq \varepsilon)$ for all $\varepsilon \geq 0$. Combining the statements yields (4.2.11). If, in addition, $\|\xi_{\ell,i}\|_{L^\infty(\Omega;V)} \leq \tau_\ell$, then the above argumentation together with Theorem 4.2.15 yield the second assertion. □

Theorems 4.7.10 and 4.7.11 are used to prove Theorem 4.2.16.

Theorem 4.7.10 ([256, Thm. 3.1], [257, Thm. 1])**.** *Let $(V,\|\cdot\|_V)$ be a $(2,\kappa)$-smooth, separable Banach space, $(Z_j)_{j\in\mathbb{N}_0}$ be a martingale adapted to the filtration $(\mathcal{F}_j)_{j\in\mathbb{N}_0} \subset \mathcal{F}$ with $Z_0 = 0$, $Z_j \in L^1(\Omega;V)$ and $\mathcal{F}_0 = \{\emptyset,\Omega\}$. Define $d_0 = 0$ and $d_j = Z_j - Z_{j-1}$ for all $j \in \mathbb{N}$. Then, for all $r \geq 0$ and every $\lambda \geq 0$, we have*

$$\mathrm{Prob}\Big(\sup_{j\in\mathbb{N}}\|Z_j\|_V \geq r\Big) \leq 2\mathrm{e}^{-\lambda r}\Big\|\prod_{j=1}^{\infty}\big(1 + \kappa\mathbb{E}[\mathrm{e}^{\lambda\|d_j\|_V} - 1 - \lambda\|d_j\|_V \mid \mathcal{F}_{j-1}]\big)\Big\|_{L^\infty(\Omega,\mathcal{F},P;\mathbb{R})}.$$

We note that the constant D in [256, p. 1680] and [257, p. 55] equals $\sqrt{\kappa}$.

Theorem 4.7.11 ([256, Thm. 3.5])**.** *If the hypotheses of Theorem 4.7.10 hold and for some $\tau > 0$, $\sum_{j=1}^{\infty}\|d_j\|_{L^\infty(\Omega;V)}^2 \leq \tau^2$, then for all $r \geq 0$, $\mathrm{Prob}(\sup_{j\in\mathbb{N}}\|Z_j\|_V \geq r) \leq 2\exp(-r^2/(2\kappa\tau^2))$.*

Lemma 4.7.12. *For all $x \geq 0$ and $\kappa \geq 1$, we have $\kappa\exp(x) + 1 - \kappa \leq \exp(\kappa x)$.*

Proof. Since $x \geq 0$ and $\kappa \geq 1$, we have $\kappa \exp(x) + 1 - \kappa = 1 + \kappa x + \kappa \sum_{k=2}^{\infty} x^k/k! \leq 1 + \kappa x + \sum_{k=2}^{\infty} (\kappa x)^k/k! = \exp(\kappa x)$. □

Proof of Theorem 4.2.16. We define $N = \sum_{\ell=1}^{L} N_\ell$, the stopped martingale $(S_j)_{j\in\mathbb{N}_0}$ as in the proof of Corollary 4.2.13, and the martingale-difference $(d_j)_{j\in\mathbb{N}_0}$ by $d_0 = 0$ and $d_j = S_j - S_{j-1}$, $j \in \mathbb{N}$. Both martingales are defined by sums of independent V-valued random variables. We omit the conditioning on their natural filtration. Now, fix ε, $\lambda \geq 0$, $L \in \mathbb{N}$, $N_\ell \in \mathbb{N}$ for $\ell = 1, \dots, L$, and fix $\ell \in \{1, \dots, L\}$. Using Lemmas 3.6.5 and 4.7.12, we find that

$$\begin{aligned} 1 + \kappa\mathbb{E}\left[\mathrm{e}^{\frac{\lambda\|\xi_{\ell,i}\|_V}{N_\ell}} - 1 - \frac{\lambda\|\xi_{\ell,i}\|_V}{N_\ell}\right] &= 1 - \kappa + \kappa\mathbb{E}\left[\mathrm{e}^{\frac{\lambda\|\xi_{\ell,i}\|_V}{N_\ell}} - \frac{\lambda\|\xi_{\ell,i}\|_V}{N_\ell}\right] \\ &\leq 1 - \kappa + \kappa\mathrm{e}^{\frac{3\lambda^2\tau_\ell^2}{4N_\ell^2}} \leq \mathrm{e}^{\frac{3\kappa\lambda^2\tau_\ell^2}{4N_\ell^2}}. \end{aligned}$$

Consequently, we obtain

$$\prod_{i=1}^{N_\ell}\left(1 + \kappa\mathbb{E}\left[\mathrm{e}^{\frac{\lambda\|\xi_{\ell,i}\|_V}{N_\ell}} - 1 - \frac{\lambda\|\xi_{\ell,i}\|_V}{N_\ell}\right]\right) \leq \prod_{i=1}^{N_\ell}\mathrm{e}^{\frac{3\kappa\lambda^2\tau_\ell^2}{4N_\ell^2}} = \mathrm{e}^{\frac{3\kappa\lambda^2}{4}\frac{\tau_\ell^2}{N_\ell}}.$$

Hence, using $d_j = 0$ for $j \geq N+1$, we get

$$\begin{aligned} \prod_{j=1}^{\infty}\left(1 + \kappa\mathbb{E}[\mathrm{e}^{\lambda\|d_j\|_V} - 1 - \lambda\|d_j\|_V]\right) &= \prod_{\ell=1}^{L}\prod_{i=1}^{N_\ell}\left(1 + \kappa\mathbb{E}\left[\mathrm{e}^{\frac{\|\lambda\xi_{\ell,i}\|_V}{N_\ell}} - 1 - \frac{\lambda\|\xi_{\ell,i}\|_V}{N_\ell}\right]\right) \\ &\leq \prod_{\ell=1}^{L}\mathrm{e}^{\frac{3\kappa\lambda^2}{4}\frac{\tau_\ell^2}{N_\ell}} = \mathrm{e}^{\frac{3\kappa\lambda^2}{4}\sum_{\ell=1}^{L}\frac{\tau_\ell^2}{N_\ell}}. \end{aligned}$$

Applying Theorem 4.7.10 gives

$$\mathrm{Prob}(\|S_N\|_V \geq \varepsilon) \leq \mathrm{Prob}\Big(\sup_{j\in\mathbb{N}}\|S_j\|_V \geq \varepsilon\Big) \leq 2\mathrm{e}^{-\lambda\varepsilon + \frac{3\kappa\lambda^2}{4}\sum_{\ell=1}^{L}\frac{\tau_\ell^2}{N_\ell}}.$$

Minimizing this bound over $\lambda > 0$, applying Lemma 3.6.6, and setting $\varepsilon = \sqrt{\kappa}r(\sum_{\ell=1}^{L}\tau_\ell^2/N_\ell)^{1/2}$ yields (4.2.10).
If, in addition, $\|\xi_{\ell,i}\|_{L^\infty(\Omega;V)} \leq \tau_\ell$ for i, $\ell = 1, 2, \dots$, then (4.7.12) and Theorem 4.7.11 yield

$$\mathrm{Prob}(\|S_N\|_V \geq \varepsilon) \leq \mathrm{Prob}\Big(\sup_{j\in\mathbb{N}}\|S_j\|_V \geq \varepsilon\Big) \leq 2\exp\Big(-\frac{\varepsilon^2}{2\kappa\sum_{\ell=1}^{L}\tau_\ell^2/N_\ell}\Big) = 2\exp(-r^2/2),$$

which implies the second assertion. □

In order to prove Theorem 4.2.18, we use the following variant of the Kahane–Khintchine inequality due to Latała and Oleszkiewicz [210].

Theorem 4.7.13 ([210, Cor. 3])**.** *If $(V, \|\cdot\|_V)$ is a separable Banach space, $(x_k) \subset V$, $g_k : \Omega \to \mathbb{R}$ are independent standard normal random variables, $\sum_{k=1}^{K} g_k x_k$ converges to Z w.p. 1, then for all $0 < p \leq q < \infty$, $\mathbb{E}[\|Z\|_V^q]^{1/q} \leq (\gamma_q/\gamma_p)(\mathbb{E}[\|Z\|_V^p])^{1/p}$, where $\gamma_r = \mathbb{E}[|g_1|^r]^{1/r}$.*

Proof of Theorem 4.2.18. We fix $\lambda \in \mathbb{R}$ and set $V = W^{s,p}(\mathcal{D})$. We estimate each addend in the series expansion $\mathbb{E}[\cosh(\lambda\|Z\|_V)] = \sum_{k=0}^{\infty}\lambda^{2k}\mathbb{E}[\|Z\|_V^{2k}]/(2k!)$. Let $g : \Omega \to \mathbb{R}$ be a standard normal random variable. For $k \in \mathbb{N}\setminus\{1\}$, we define $\gamma_k = \mathbb{E}[|g|^k]^{1/k}$. We have $\gamma_k \geq 1$, and

Lemma 4.7.7 implies $\gamma_k^k \le 2(k/\mathrm{e})^{k/2}$. Furthermore $\gamma_{2k}^{2k} = (2k)!/(2^k k!)$ for each $k \in \mathbb{N}$ (see, e.g., [57, p. 227]).
Fix $k \in \mathbb{N}$ with $2k \ge p$. Theorem 4.7.13 (applied with $g_k = \xi_k/\tau_k$ and $x_k = \tau_k\phi_k$ for $\tau_k > 0$ and $1 \le k \le K$, and $x_k = 0$ for $k \ge K+1$) implies $\mathbb{E}[\|Z\|_V^{2k}] \le (\gamma_{2k}/\gamma_p)^{2k}(\mathbb{E}[\|Z\|_V^p])^{2k/p}$. Proposition 4.2.14 yields $\mathbb{E}[\|Z\|_V^p] = \gamma_p^p T_K^p$, where T_K is defined in (4.2.8). Combining both inequalities with $\gamma_{2k}^{2k} = (2k)!/(2^k k!)$ ensures $\mathbb{E}[\|Z\|_V^{2k}] \le (\gamma_{2k}/\gamma_p)^{2k}\gamma_p^{2k}T_K^{2k} = (2k)!/(2^k k!)T_K^{2k}$.
Fix $k \in \mathbb{N}$ with $2k < p$. Jensen's inequality implies $\mathbb{E}[\|Z\|_V^{2k}] \le \mathbb{E}[\|Z\|_V^p]^{2k/p}$. Combined with $\mathbb{E}[\|Z\|_V^p] = \gamma_p^p T_K^p$ and $\gamma_p^{2k} \le \gamma_p^p$, we find that $\mathbb{E}[\|Z\|_V^{2k}] \le \gamma_p^{2k}T_K^{2k} \le \gamma_p^p T_K^{2k}$.
Putting together the pieces, and using $(2k)! \ge 2^k k!$ valid for $k \in \mathbb{N}$, we find that

$$\mathbb{E}[\cosh(\lambda\|Z\|_V)] \le 1 + \gamma_p^p \sum_{k=1}^{\lceil p/2\rceil - 1} \frac{\lambda^{2k}T_K^{2k}}{(2k)!} + \sum_{k=\lceil p/2\rceil}^{\infty} \frac{\lambda^{2k}T_K^{2k}}{2^k k!} \le \gamma_p^p \exp(\lambda^2 T_K^2/2).$$

It must yet be shown that $\mathrm{Prob}(\|Z\|_V \ge r) \le 2\gamma_p^p \exp(-r^2/(2T_K^2))$ for $r > 0$. Using $\exp(x) \le 2\cosh(x)$, valid for all $x \in \mathbb{R}^n$, and Markov's inequality, we obtain for each $\lambda > 0$,

$$\mathrm{Prob}(\|Z\|_V \ge r) \le 2\mathrm{e}^{-\lambda r}\mathbb{E}[\cosh(\lambda\|Z\|_V)] \le 2\gamma_p^p \mathrm{e}^{\lambda^2 T_K^2/2 - \lambda r}.$$

Minimizing the right-hand side over $\lambda > 0$ yields the tail bound. □

4.7.5 Proof of a Technical Lemma

We recall that $\lceil x\rceil_{\mathbb{N}} \in \mathbb{N}$ is the smallest number such that $x \le \lceil x\rceil_{\mathbb{N}}$ for $x \in \mathbb{R}$, and that $0 \notin \mathbb{N}$; see p. vii and section 4.2.

Proof of Lemma 4.3.3. The proof is based on that of [81, Thm. 1]. Since $m_L \le c_\alpha s^{-(L-1)\alpha}h_1^\alpha$, we have $m_L \le \epsilon$. Because $L = \lceil \alpha^{-1}\log_s(c_\alpha\epsilon^{-1}h_1^\alpha) + 1\rceil_{\mathbb{N}}$, we have $L = 1$ if and only if $c_\alpha\epsilon^{-1}h_1^\alpha \le 1$. If $L = 1$, then we choose $N_1 = \lceil \epsilon^{-2}\eta^{-1}c_\beta h_1^\beta\rceil_{\mathbb{N}}$, which ensures $\tau_1^2/N_1 \le \epsilon^2\eta$. For the remainder of the proof, let $L \ge 2$. We obtain

$$\sum_{\ell=1}^{L} s^{\gamma(\ell-1)} = \sum_{\ell=0}^{L-1} s^{\gamma\ell} < \frac{s^{\gamma(L-1)}}{1-s^{-\gamma}} \le \frac{s^\gamma c_\alpha^{\gamma/\alpha}}{1-s^{-\gamma}}\epsilon^{-\gamma/\alpha}. \tag{4.7.13}$$

Similar to the choices made in the proof of [81, Thm. 1], we define

$$N_\ell = \begin{cases} \lceil \epsilon^{-2}\eta^{-1}c_\beta(1-s^{-(\beta-\gamma)/2})^{-1}s^{-(\beta+\gamma)(\ell-1)/2}h_1^\beta\rceil_{\mathbb{N}} & \text{if } \beta > \gamma, \\ \lceil \epsilon^{-2}\eta^{-1}Lc_\beta s^{-\beta(\ell-1)}h_1^\beta\rceil_{\mathbb{N}} & \text{if } \beta = \gamma, \\ \lceil \epsilon^{-2}\eta^{-1}c_\beta s^{(\gamma-\beta)(L-1)/2}(1-s^{-(\gamma-\beta)/2})^{-1}s^{-(\beta+\gamma)(\ell-1)/2}h_1^\beta\rceil_{\mathbb{N}} & \text{if } \beta < \gamma. \end{cases} \tag{4.7.14}$$

Owing to (4.7.14) and the definition of $\lceil\cdot\rceil_{\mathbb{N}}$ (see section 4.2), we have $N_\ell \le (1/s)^{(\beta+\gamma)/2}N_{\ell-1} + 1$ and $N_\ell \le (1/s)^{(\beta+\gamma)(\ell-1)/2}N_1 + 1$.
If $\beta = \gamma$, then we obtain, with $h_\ell^\beta \le s^{-\beta(\ell-1)}h_1^\beta$ and (4.7.14),

$$\sum_{\ell=1}^{L}\frac{\tau_\ell^2}{N_\ell} \le \sum_{\ell=1}^{L}\frac{c_\beta h_\ell^\beta}{N_\ell} = \sum_{N_\ell=1} c_\beta h_\ell^\beta + \sum_{N_\ell>1}\frac{c_\beta h_\ell^\beta}{N_\ell} \le \epsilon^2\eta\sum_{\ell=1}^{L}\frac{c_\beta s^{-(\ell-1)\beta}h_1^\beta}{c_\beta s^{-(\ell-1)\beta}h_1^\beta L} = \epsilon^2\eta.$$

Combining $h_\ell^{-\gamma} \le s^{\gamma(\ell-1)}h_1^{-\gamma}$ and (4.7.14), we obtain

$$\sum_{\ell=1}^{L} N_\ell h_\ell^{-\gamma} = \sum_{N_\ell=1} h_\ell^{-\gamma} + \sum_{N_\ell>1} h_\ell^{-\gamma} \le \epsilon^{-2}\eta^{-1}L^2c_\beta + \sum_{\ell=1}^{L} h_\ell^{-\gamma} = \epsilon^{-2}\eta^{-1}L^2c_\beta + \sum_{\ell=1}^{L} s^{\gamma(\ell-1)}h_1^{-\gamma}.$$

We have $L^2 \leq (\alpha^{-1}\log_s(c_\alpha \epsilon^{-1} h_1^\alpha)+2)^2$. Combined with (4.7.13), we obtain the bound in (4.3.6) for $\beta = \gamma$.
If $\beta > \gamma$, then (4.7.14) yields

$$\sum_{\ell=1}^{L} \frac{\tau_\ell^2}{N_\ell} \leq \epsilon^2 \eta (1 - s^{-(\beta-\gamma)/2}) \sum_{\ell=1}^{L} \frac{s^{-\beta(\ell-1)}}{s^{-(\beta+\gamma)(\ell-1)/2}} \leq \epsilon^2 \eta (1 - s^{-(\beta-\gamma)/2}) \sum_{\ell=0}^{\infty} s^{-(\beta-\gamma)\ell/2} \leq \epsilon^2 \eta.$$

Moreover, using (4.7.14), we find that

$$\sum_{\ell=1}^{L} N_\ell h_\ell^{-\gamma} \leq \epsilon^{-2}\eta^{-1} c_\beta (1 - s^{-(\beta-\gamma)/2})^{-1} \sum_{\ell=1}^{L} s^{-(\beta-\gamma)(\ell-1)/2} h_1^{\beta-\gamma} + \sum_{\ell=1}^{L} s^{\gamma(\ell-1)} h_1^{-\gamma}.$$

Combined with (4.7.13), we obtain the bound in (4.3.6) for $\beta > \gamma$.
If $\beta < \gamma$, then

$$\sum_{\ell=1}^{L} \frac{\tau_\ell^2}{N_\ell} \leq \epsilon^2 \eta (1 - s^{-(\gamma-\beta)/2}) s^{-(\gamma-\beta)(L-1)/2} \sum_{\ell=0}^{\infty} s^{-(\gamma-\beta)\ell/2} \leq \epsilon^2 \eta.$$

Furthermore,

$$\sum_{\ell=1}^{L} N_\ell h_\ell^{-\gamma} \leq \epsilon^{-2}\eta^{-1} c_\beta s^{(\gamma-\beta)(L-1)} (1 - s^{-(\gamma-\beta)/2})^{-2} h_1^{\beta-\gamma} + \sum_{\ell=1}^{L} s^{\gamma(\ell-1)} h_1^{-\gamma}.$$

Since $L \geq 2$, we have $L - 1 \leq \alpha^{-1}\log_s(c_\alpha \epsilon^{-1} h_1^\alpha) + 1$. Hence,

$$s^{(\gamma-\beta)(L-1)} h_1^{\beta-\gamma} = s^{\gamma-\beta} c_\alpha^{(\gamma-\beta)/\alpha} \epsilon^{-(\gamma-\beta)/\alpha} h_1^{\gamma-\beta} h_1^{\beta-\gamma} = s^{\gamma-\beta} c_\alpha^{(\gamma-\beta)/\alpha} \epsilon^{-(\gamma-\beta)/\alpha}.$$

Combined with (4.7.13), we obtain the bound in (4.3.6) for $\beta < \gamma$. □

Acknowledgements

I am very grateful to have had the privilege of conducting research with my advisor, Professor Dr. Michael Ulbrich, over the past three years. As an advisor, he fosters creativity and independent thinking, while being always ready to provide assistance when needed. I thank Professor Dr. Michael Ulbrich for his guidance and constructive suggestions on writing papers. Moreover, I thank Professor Dr. Karl Kunisch for being my co-advisor and for our discussions, which have enriched my perspective on research. I appreciate the valuable feedback and suggestions for improvement from Professor Dr. Michael Ulbrich and Professor Dr. Karl Kunisch on the earlier drafts of my dissertation.

I thank my examiners, Professor Dr. Michael Ulbrich, Professor Dr. Karl Kunisch, and Professor Dr. Alexander Shapiro for reviewing and evaluating my dissertation, and Professor Dr. Rainer Callies for serving as the head of my dissertation committee.

I acknowledge the financial and educational support from the International Research Training Group IGDK Munich — Graz *Optimization and Numerical Analysis for Partial Differential Equations with Nonsmooth Structures* (project number 188264188/GRK1754), funded by the German Science Foundation (DFG) and the Austrian Science Fund (FWF).

I thank Professor Dr. Michael Ulbrich and Dr. Christian Ludwig for explaining how to implement Julia interfaces for Fortran(77) codes. I appreciate that Dr. Christian Ludwig has allowed me to reuse large parts of his ODEInterface.jl-code.

I appreciate the valuable feedback and suggestions for improvements from Niklas Behringer, Dr. Johannes Haubner, Gernot Holler, Fabian Schaipp, and Julia Wachter on parts of the earlier drafts of my dissertation. I have enjoyed and benefited from the scientific discussions on different subjects with Niklas Behringer, Dr. Sebastian Garreis, Dr. Johannes Haubner, Lukas Hertlein, Gernot Holler, Franziska Neumann, Fabian Schaipp, Julia Wachter, and Dr. Daniel Walter.

I thank my parents, Lisa and Josef, my siblings, Katharina, Sebastian, and Jakob, and my friends.

Bibliography

[1] R. A. Adams, *Sobolev Spaces*, Academic Press, New York, NY, 1975. (Cited on pp. ix, 51, 52, 86, 90, 91, 92, 114, 119, and 133.)

[2] A. Alexanderian, N. Petra, G. Stadler, and O. Ghattas, *Mean-Variance Risk-Averse Optimal Control of Systems Governed by PDEs with Random Parameter Fields Using Quadratic Approximations*, SIAM/ASA J. Uncertainty Quantification, 5 (2017), pp. 1166–1192, https://doi.org/10.1137/16M106306X. (Cited on p. 41.)

[3] A. A. Ali, E. Ullmann, and M. Hinze, *Multilevel Monte Carlo Analysis for Optimal Control of Elliptic PDEs with Random Coefficients*, SIAM/ASA J. Uncertainty Quantification, 5 (2017), pp. 466–492, https://doi.org/10.1137/16M109870X. (Cited on pp. 86, 89, 90, and 93.)

[4] A. Alla, M. Hinze, P. Kolvenbach, O. Lass, and S. Ulbrich, *A certified model reduction approach for robust parameter optimization with PDE constraints*, Adv. Comput. Math., 45 (2019), pp. 1221–1250, https://doi.org/10.1007/s10444-018-9653-1. (Cited on pp. 10, 41, and 42.)

[5] M. S. Alnæs, *UFL: a finite element form language*, in Automated Solution of Differential Equations by the Finite Element Method: The FEniCS Book, A. Logg, K.-A. Mardal, and G. Wells, eds., Springer, Berlin, 2012, pp. 303–338, https://doi.org/10.1007/978-3-642-23099-8_17. (Cited on pp. 40 and 54.)

[6] M. S. Alnæs, J. Blechta, J. Hake, A. Johansson, B. Kehlet, A. Logg, C. Richardson, J. Ring, M. E. Rognes, and G. N. Wells, *The FEniCS Project Version 1.5*, Arch. Numer. Software, 3 (2015), pp. 9–23, https://doi.org/10.11588/ans.2015.100.20553. (Cited on pp. 40, 54, and 124.)

[7] M. S. Alnæs, A. Logg, K. B. Ølgaard, M. E. Rognes, and G. N. Wells, *Unified Form Language: A Domain-specific Language for Weak Formulations of Partial Differential Equations*, ACM Trans. Math. Softw., 40 (2014), pp. 9:1–9:37, https://doi.org/10.1145/2566630. (Cited on pp. 40 and 54.)

[8] R. Andreani, G. Haeser, and J. M. Martínez, *On sequential optimality conditions for smooth constrained optimization*, Optimization, 60 (2011), pp. 627–641, https://doi.org/10.1080/02331930903578700. (Cited on p. 12.)

[9] R. G. Antonini, *Subgaussian random variables in Hilbert spaces*, Rend. Semin. Mat. Univ. Padova, 98 (1997), pp. 89–99, http://www.numdam.org/item/RSMUP_1997__98__89_0. (Cited on p. 110.)

[10] Z. Artstein and R. J. B. Wets, *Consistency of minimizers and the SLLN for stochastic programs*, J. Convex Anal., 2 (1995), pp. 1–17. (Cited on p. 72.)

[11] J.-P. Aubin and H. Frankowska, *Set-Valued Analysis*, Mod. Birkhäuser Class., Springer, Boston, MA, 2009, https://doi.org/10.1007/978-0-8176-4848-0. (Cited on pp. 74, 85, and 102.)

[12] I. Babuška, F. Nobile, and R. Tempone, *A stochastic collocation method for elliptic partial differential equations with random input data*, SIAM J. Numer. Anal., 45 (2007), pp. 1005–1034, https://doi.org/10.1137/050645142. (Cited on pp. 86, 121, 122, and 123.)

[13] K. Ball, E. A. Carlen, and E. H. Lieb, *Sharp uniform convexity and smoothness inequalities for trace norms*, Invent. Math., 115 (1994), pp. 463–482, https://doi.org/10.1007/bf01231769. (Cited on pp. 111, 112, 129, and 130.)

[14] V. Barbu and T. Precupanu, *Convexity and Optimization in Banach Spaces*, Springer Monogr. Math., Springer, Dordrecht, 4th ed., 2012, https://doi.org/10.1007/978-94-007-2247-7. (Cited on pp. 112 and 132.)

[15] A. Barth, A. Lang, and Ch. Schwab, *Multilevel Monte Carlo method for parabolic stochastic partial differential equations*, BIT Numer. Math., 53 (2013), pp. 3–27, https://doi.org/10.1007/s10543-012-0401-5. (Cited on p. 115.)

[16] A. Barth, Ch. Schwab, and N. Zollinger, *Multi-level Monte Carlo Finite Element method for elliptic PDEs with stochastic coefficients*, Numer. Math., 119 (2011), pp. 123–161, https://doi.org/10.1007/s00211-011-0377-0. (Cited on pp. 108, 115, 120, 122, 123, 124, and 125.)

[17] P. L. Bartlett, S. R. Kulkarni, and S. E. Posner, *Covering numbers for real-valued function classes*, IEEE Trans. Inf. Theory, 43 (1997), pp. 1721–1724, https://doi.org/10.1109/18.623181. (Cited on p. 71.)

[18] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books Math., Springer, New York, NY, 2011, https://doi.org/10.1007/978-1-4419-9467-7. (Cited on p. 75.)

[19] M. Ben Alaya and A. Kebaier, *Central limit theorem for the multilevel Monte Carlo Euler method*, Ann. Appl. Probab., 25 (2015), pp. 211–234, https://doi.org/10.1214/13-AAP993. (Cited on p. 109.)

[20] A. Ben-Tal and D. den Hertog, *Hidden conic quadratic representation of some nonconvex quadratic optimization problems*, Math. Program., 143 (2014), pp. 1–29, https://doi.org/10.1007/s10107-013-0710-8. (Cited on p. 10.)

[21] A. Ben-Tal, D. den Hertog, A. De Waegenaere, B. Melenberg, and G. Rennen, *Robust solutions of optimization problems affected by uncertain probabilities*, Manag. Sci., 59 (2013), pp. 341–357, https://doi.org/10.1287/mnsc.1120.1641. (Cited on pp. 2 and 9.)

[22] A. Ben-Tal, D. den Hertog, and J.-P. Vial, *Deriving robust counterparts of nonlinear uncertain inequalities*, Math. Program., 149 (2015), pp. 265–299, https://doi.org/10.1007/s10107-014-0750-8. (Cited on p. 10.)

[23] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski, *Robust Optimization*, Princeton Ser. Appl. Math., Princeton University Press, Princeton, NJ, 2009. (Cited on pp. 2, 10, 27, 37, and 40.)

[24] A. Ben-Tal and A. Nemirovski, *Robust solutions of uncertain linear programs*, Oper. Res. Lett., 25 (1999), pp. 1–13, https://doi.org/10.1016/S0167-6377(99)00016-4. (Cited on p. 2.)

[25] A. Ben-Tal and A. Nemirovski, *Robust solutions of linear programming problems contaminated with uncertain data*, Math. Program., 88 (2000), pp. 411–424, https://doi.org/10.1007/PL00011380. (Cited on pp. 5 and 34.)

[26] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization*, MPS-SIAM Ser. Optim. 2, SIAM, Philadelphia, PA, 2001, https://doi.org/10.1137/1.9780898718829. (Cited on pp. 9, 10, 36, and 37.)

[27] A. Ben-Tal and M. Teboulle, *Expected utility, penalty functions, and duality in stochastic nonlinear programming*, Management Sci., 32 (1986), pp. 1445–1466, https://doi.org/10.1287/mnsc.32.11.1445. (Cited on p. 96.)

[28] A. Ben-Tal and M. Teboulle, *Hidden convexity in some nonconvex quadratically constrained quadratic programming*, Math. Program., 72 (1996), pp. 51–63, https://doi.org/10.1007/BF02592331. (Cited on pp. 9 and 15.)

[29] P. Benner, S. Dolgov, A. Onwunta, and M. Stoll, *Low-rank solvers for unsteady Stokes–Brinkman optimal control problem with random data*, Comput. Methods Appl. Mech. Engrg., 304 (2016), pp. 26–54, https://doi.org/10.1016/j.cma.2016.02.004. (Cited on pp. 41, 73, and 101.)

[30] D. Bertsimas, D. B. Brown, and C. Caramanis, *Theory and Applications of Robust Optimization*, SIAM Rev., 53 (2011), pp. 464–501, https://doi.org/10.1137/080734510. (Cited on pp. 2 and 10.)

[31] D. Bertsimas, X. V. Doan, K. Natarajan, and C.-P. Teo, *Models for minimax stochastic linear optimization problems with risk aversion*, Math. Oper. Res., 35 (2010), pp. 580–602, https://doi.org/10.1287/moor.1100.0445. (Cited on pp. 10 and 27.)

[32] D. Bertsimas, O. Nohadani, and K. M. Teo, *Nonconvex Robust Optimization for Problems with Constraints*, INFORMS J. Comput., 22 (2010), pp. 44–58, https://doi.org/10.1287/ijoc.1090.0319. (Cited on pp. 10, 11, and 27.)

[33] D. Bertsimas, O. Nohadani, and K. M. Teo, *Robust Optimization for Unconstrained Simulation-Based Problems*, Oper. Res., 58 (2010), pp. 161–178, https://doi.org/10.1287/opre.1090.0715. (Cited on pp. 10 and 11.)

[34] D. Bertsimas and J. Sethuraman, *Moment Problems and Semidefinite Optimization*, in Handbook of Semidefinite Programming: Theory, Algorithms, and Applications, H. Wolkowicz, R. Saigal, and L. Vandenberghe, eds., Springer, Boston, MA, 2000, pp. 469–509, https://doi.org/10.1007/978-1-4615-4381-7_16. (Cited on p. 9.)

[35] J. Bezanson, A. Edelman, S. Karpinski, and V. Shah, *Julia: A Fresh Approach to Numerical Computing*, SIAM Rev., 59 (2017), pp. 65–98, https://doi.org/10.1137/141000671. (Cited on pp. 28 and 37.)

[36] A. T. Bharucha-Reid, *Random Integral Equations*, Math. Sci. Eng. 96, Academic Press, New York, 1972. (Cited on pp. 78, 84, 85, 88, and 122.)

[37] C. Bierig and A. Chernov, *Convergence analysis of multilevel Monte Carlo variance estimators and application for random obstacle problems*, Numer. Math., 130 (2014), pp. 579–613, https://doi.org/10.1007/s00211-014-0676-3. (Cited on pp. 1, 4, 105, 106, 108, 109, 113, 114, 115, 116, 119, 125, and 129.)

[38] C. Bierig and A. Chernov, *Estimation of arbitrary order central statistical moments by the multilevel Monte Carlo method*, Stoch. PDE: Anal. Comp., 4 (2016), pp. 3–40, https://doi.org/10.1007/s40072-015-0063-9. (Cited on p. 108.)

[39] P. Billingsley, *Probability and Measure*, Wiley Ser. Probab. Stat., John Wiley & Sons, Hoboken, NJ, 2012. Anniversary Edition. (Cited on pp. ix and 45.)

[40] M. Š. Birman and M. Z. Solomjak, *Piecewise-polynomial approximations of functions of the classes W_p^α*, Math. USSR Sb., 2 (1967), pp. 295–317, https://doi.org/10.1070/sm1967v002n03abeh002343. (Cited on pp. 72 and 101.)

[41] M. Š. Birman and M. Z. Solomjak, *Quantitative Analysis in Sobolev Imbedding Theorems and Applications to Spectral Theory*, Amer. Math. Soc. Transl. Ser. 2, 114, American Mathematical Society, Providence, R.I., 1980. Translated from the Russian by F. A. Cezus. (Cited on pp. 72 and 101.)

[42] Å. Björck, *Numerical Methods for Least Squares Problems*, Other Titles in Applied Mathematics, SIAM, Philadelphia, PA, 1996, https://doi.org/10.1137/1.9781611971484. (Cited on p. 9.)

[43] J. Blanchet and K. Murthy, *Quantifying distributional model risk via optimal transport*, Math. Oper. Res., 44 (2019), pp. 565–600, https://doi.org/10.1287/moor.2018.0936. (Cited on p. 2.)

[44] V. I. Bogachev, *Measure Theory*, Springer, Berlin, 2007, https://doi.org/10.1007/978-3-540-34514-5. (Cited on pp. 73, 79, 88, 106, 115, and 122.)

[45] V. I. Bogachev, *Weak Convergence of Measures*, Math. Surveys Monogr. 234, American Mathematical Society, Providence, RI, 2018. (Cited on pp. 44, 46, and 47.)

[46] J. F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer Ser. Oper. Res., Springer, New York, NY, 2000, https://doi.org/10.1007/978-1-4612-1394-9. (Cited on pp. viii, ix, 21, 44, 46, 49, 51, 63, 74, 75, 78, 81, 86, 88, 90, 92, 93, 98, 101, 102, and 131.)

[47] B. Borchers, *CSDP, A C library for semidefinite programming*, Optim. Methods Softw., 11 (1999), pp. 613–623, https://doi.org/10.1080/10556789908805765. (Cited on p. 38.)

[48] B. Borchers and J. G. Young, *Implementation of a primal-dual method for SDP on a shared memory parallel architecture*, Comput. Optim. Appl., 37 (2007), pp. 355–369, https://doi.org/10.1007/s10589-007-9030-3. (Cited on p. 38.)

[49] F. Bornemann, *Numerical Linear Algebra*, Springer Undergrad. Math. Ser., Springer, Cham, 2018, https://doi.org/10.1007/978-3-319-74222-9. (Cited on p. 5.)

[50] J. M. Borwein and J. D. Vanderwerff, *Convex functions: Constructions, Characterizations and Counterexamples*, Encyclopedia Math. Appl. 109, Cambridge University Press, Cambridge, 2010, https://doi.org/10.1017/CBO9781139087322. (Cited on pp. 111, 131, and 132.)

[51] A. Borzì and V. Schulz, *Computational Optimization of Systems Governed by Partial Differential Equations*, Comput. Sci. Eng. 8, SIAM, Philadelphia, 2011, https://doi.org/10.1137/1.9781611972054. (Cited on p. 73.)

[52] S. Boucheron, G. Lugosi, and P. Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence*, Oxford University Press, Oxford, 2013. (Cited on p. 113.)

[53] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, 2004, https://doi.org/10.1017/CBO9780511804441. (Cited on pp. 9 and 37.)

[54] S. C. Brenner and L. R. Scott, *The Mathematical Theory of Finite Element Methods*, Texts Appl. Math. 15, Springer, New York, 3rd ed., 2008, https://doi.org/10.1007/978-0-387-75934-0. (Cited on pp. 76, 90, 119, 121, 124, and 125.)

[55] P. Bühlmann and S. van de Geer, *Statistics for High-Dimensional Data: Methods, Theory and Applications*, Springer Ser. Statist., Springer, Berlin, 2011, https://doi.org/10.1007/978-3-642-20192-9. (Cited on p. 113.)

[56] V. Buldygin and S. Solntsev, *Asymptotic Behaviour of Linearly Transformed Sums of Random Variables*, Math. Appl. 416, Springer, Dordrecht, 1997, https://doi.org/10.1007/978-94-011-5568-7. (Cited on pp. 108 and 111.)

[57] V. V. Buldygin and Yu. V. Kozachenko, *Metric Characterization of Random Variables and Random Processes*, Transl. Math. Monogr. 188, American Mathematical Society, Providence, RI, 2000. (Cited on pp. ix, 5, 8, 35, 36, 45, 46, 77, 87, 101, 106, 107, 108, 110, 111, 123, 133, and 136.)

[58] J. V. Burke and T. Hoheisel, *Epi-convergent Smoothing with Applications to Convex Composite Functions*, SIAM J. Optim., 23 (2013), pp. 1457–1479, https://doi.org/10.1137/120889812. (Cited on pp. 10 and 12.)

[59] J. V. Burke and T. Hoheisel, *Epi-convergence Properties of Smoothing by Infimal Convolution*, Set-Valued Var. Anal., 25 (2017), pp. 1–23, https://doi.org/10.1007/s11228-016-0362-y. (Cited on pp. 10 and 12.)

[60] J. V. Burke, T. Hoheisel, and C. Kanzow, *Gradient Consistency for Integral-convolution Smoothing Functions*, Set-Valued Var. Anal., 21 (2013), pp. 359–376, https://doi.org/10.1007/s11228-013-0235-6. (Cited on p. 12.)

[61] J. V. Burke, A. S. Lewis, and M. L. Overton, *A Robust Gradient Sampling Algorithm for Nonsmooth, Nonconvex Optimization*, SIAM J. Optim., 15 (2005), pp. 751–779, https://doi.org/10.1137/030601296. (Cited on p. 9.)

[62] C. Büskens and R. Griesse, *Parametric sensitivity analysis of perturbed PDE optimal control problems with state and control constraints*, J. Optim. Theory Appl., 131 (2006), pp. 17–35, https://doi.org/10.1007/s10957-006-9122-8. (Cited on p. 56.)

[63] G. Calafiore and M. Campi, *Uncertain convex programs: randomized solutions and confidence levels*, Math. Program., 102 (2005), pp. 25–46, https://doi.org/10.1007/s10107-003-0499-y. (Cited on p. 2.)

[64] C. Carstensen, *Quasi-interpolation and a posteriori error analysis in finite element methods*, ESAIM Math. Model. Numer. Anal., 33 (1999), pp. 1187–1202, https://doi.org/10.1051/m2an:1999140. (Cited on p. 91.)

[65] E. Casas and F. Tröltzsch, *Error estimates for linear-quadratic elliptic control problems*, in Analysis and Optimization of Differential Systems, V. Barbu, I. Lasiecka,

D. Tiba, and C. Varsan, eds., Congress Ser. 121, Boston, MA, 2003, Springer, pp. 89–100, https://doi.org/10.1007/978-0-387-35690-7_10. (Cited on pp. 70 and 72.)

[66] C. Castaing and M. Valadier, *Convex Analysis and Measurable Multifunctions*, Lecture Notes in Math. 580, Springer, Berlin, 1977, https://doi.org/10.1007/bfb0087685. (Cited on pp. 52, 74, 85, and 102.)

[67] O. Catoni, *Challenging the empirical mean and empirical variance: A deviation study*, Ann. Inst. Henri Poincaré Probab. Stat., 48 (2012), pp. 1148–1185, https://doi.org/10.1214/11-AIHP454. (Cited on p. 129.)

[68] T. F. Chan, G. H. Golub, and R. J. Leveque, *Algorithms for Computing the Sample Variance: Analysis and Recommendations*, Amer. Statist., 37 (1983), pp. 242–247, https://doi.org/10.1080/00031305.1983.10483115. (Cited on p. 124.)

[69] J. Charrier, *Strong and weak error estimates for elliptic partial differential equations with random coefficients*, SIAM J. Numer. Anal., 50 (2012), pp. 216–246, https://doi.org/10.1137/100800531. (Cited on pp. 86, 90, 120, 123, and 129.)

[70] J. Charrier, R. Scheichl, and A. Teckentrup, *Finite Element Error Analysis of Elliptic PDEs with Random Coefficients and Its Application to Multilevel Monte Carlo Methods*, SIAM J. Numer. Anal., 51 (2013), pp. 322–352, https://doi.org/10.1137/110853054. (Cited on pp. 93, 109, 120, 121, 123, and 129.)

[71] P. Chen, U. Villa, and O. Ghattas, *Taylor approximation for PDE-constrained optimization under uncertainty: Application to turbulent jet flow*, Proc. Appl. Math. Mech., 18 (2018), p. e201800466, https://doi.org/10.1002/pamm.201800466. (Cited on p. 41.)

[72] X. Chen, *Smoothing methods for nonsmooth, nonconvex minimization*, Math. Program., 134 (2012), pp. 71–99, https://doi.org/10.1007/s10107-012-0569-0. (Cited on pp. 9, 10, 11, and 12.)

[73] X. Chen, Z. Nashed, and L. Qi, *Smoothing methods and semismooth methods for nondifferentiable operator equations*, SIAM J. Numer. Anal., 38 (2000), pp. 1200–1216, https://doi.org/10.1137/S0036142999356719. (Cited on pp. 11, 12, and 42.)

[74] X. Chen, H. Qi, L. Qi, and K.-L. Teo, *Smooth Convex Approximation to the Maximum Eigenvalue Function*, J. Global Optim., 30 (2004), pp. 253–270, https://doi.org/10.1007/s10898-004-8271-2. (Cited on p. 17.)

[75] Z. Chen, M. Sim, and H. Xu, *Distributionally Robust Optimization with Infinitely Constrained Ambiguity Sets*, Oper. Res., 67 (2019), pp. 1328–1344, https://doi.org/10.1287/opre.2018.1799. (Cited on pp. 2, 8, 9, 40, and 62.)

[76] H. Chernoff, *A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations*, Ann. Math. Statist., 23 (1952), pp. 493–507, https://doi.org/10.1214/aoms/1177729330. (Cited on p. 103.)

[77] C. Chidume, *Geometric Properties of Banach Spaces and Nonlinear Iterations*, Lecture Notes in Math. 1965, Springer, London, 2009, https://doi.org/10.1007/978-1-84882-190-3. (Cited on p. 112.)

[78] F. H. Clarke, *Generalized gradients and applications*, Trans. Amer. Math. Soc., 205 (1975), pp. 247–262, https://doi.org/10.1090/S0002-9947-1975-0367131-6. (Cited on pp. 15, 24, and 27.)

[79] F. H. Clarke, *Optimization and Nonsmooth Analysis*, Classics Appl. Math. 5, SIAM, Philadelphia, PA, 1990, https://doi.org/10.1137/1.9781611971309. (Cited on pp. ix, 15, 22, 27, and 29.)

[80] C. Clason and B. Kaltenbacher, *Optimal control and inverse problems*, Inverse Probl., 36 (2020), p. 060301, https://doi.org/10.1088/1361-6420/ab8485. (Cited on p. 42.)

[81] K. A. Cliffe, M. B. Giles, R. Scheichl, and A. L. Teckentrup, *Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients*, Comput. Vis. Sci., 14 (2011), pp. 3–15, https://doi.org/10.1007/s00791-011-0160-x. (Cited on pp. 116 and 136.)

[82] N. Collier, A.-L. Haji-Ali, F. Nobile, E. von Schwerin, and R. Tempone, *A continuation multilevel Monte Carlo algorithm*, BIT Numer. Math., 55 (2015), pp. 399–432, https://doi.org/10.1007/s10543-014-0511-3. (Cited on p. 108.)

[83] A. R. Conn and L. N. Vicente, *Bilevel derivative-free optimization and its application to robust optimization*, Optim. Methods Softw., 27 (2012), pp. 561–577, https://doi.org/10.1080/10556788.2010.547579. (Cited on p. 10.)

[84] S. Conti, H. Held, M. Pach, M. Rumpf, and R. Schultz, *Shape Optimization Under Uncertainty—A Stochastic Programming Perspective*, SIAM J. Optim., 19 (2009), pp. 1610–1632, https://doi.org/10.1137/070702059. (Cited on p. 41.)

[85] S. Conti, H. Held, M. Pach, M. Rumpf, and R. Schultz, *Risk Averse Shape Optimization*, SIAM J. Control Optim., 49 (2011), pp. 927–947, https://doi.org/10.1137/090754315. (Cited on p. 41.)

[86] C. Czado and T. Schmidt, *Mathematische Statistik*, Statistik und ihre Anwendungen, Springer, Berlin, 2011, https://doi.org/10.1007/978-3-642-17261-8. (Cited on p. 76.)

[87] M. Dashti and A. M. Stuart, *The Bayesian Approach to Inverse Problems*, in Handbook of Uncertainty Quantification, R. Ghanem, D. Higdon, and H. Owhadi, eds., Springer, Cham, 2016, pp. 1–118, https://doi.org/10.1007/978-3-319-11259-6_7-1. (Cited on pp. 108 and 111.)

[88] Th. Daun and S. Heinrich, *Complexity of Banach Space Valued and Parametric Integration*, in Monte Carlo and Quasi-Monte Carlo Methods 2012, J. Dick, F. Y. Kuo, G. W. Peters, and I. H. Sloan, eds., Springer Proc. Math. Stat. 56, Berlin, 2013, Springer, pp. 297–316, https://doi.org/10.1007/978-3-642-41095-6_12. (Cited on p. 109.)

[89] Th. Daun and S. Heinrich, *Complexity of parametric integration in various smoothness classes*, J. Complexity, 30 (2014), pp. 750–766, https://doi.org/10.1016/j.jco.2014.04.002. (Cited on p. 109.)

[90] F. de Gournay, G. Allaire, and F. Jouve, *Shape and topology optimization of the robust compliance via the level set method*, ESAIM Control. Optim. Calc. Var., 14 (2008), pp. 43–70, https://doi.org/10.1051/cocv:2007048. (Cited on p. 41.)

[91] J. C. De los Reyes, *Numerical PDE-Constrained Optimization*, SpringerBriefs Optim., Springer, Cham, 2015, https://doi.org/10.1007/978-3-319-13395-9. (Cited on p. 84.)

[92] J. C. De los Reyes and K. Kunisch, *A comparison of algorithms for control constrained optimal control of the Burgers equation*, Calcolo, 41 (2004), pp. 203–225, https://doi.org/10.1007/s10092-004-0092-7. (Cited on pp. 3 and 62.)

[93] J. C. De los Reyes, C. Meyer, and B. Vexler, *Finite element error analysis for state-constrained optimal control of the Stokes equations*, Control Cybernet., 37 (2008), pp. 251–284, http://control.ibspan.waw.pl:3000/contents/export?filename=2008-2-01_reyes_et_al.pdf. (Cited on pp. 90, 91, and 119.)

[94] E. Delage and Y. Ye, *Distributionally Robust Optimization Under Moment Uncertainty with Application to Data-Driven Problems*, Oper. Res., 58 (2010), pp. 595–612, https://doi.org/10.1287/opre.1090.0741. (Cited on pp. 1, 2, 5, 8, 9, 10, 28, 36, 40, 41, and 62.)

[95] R. Deville, G. Godefroy, and V. Zizler, *Smoothness and Renormings in Banach Spaces*, Pitman Monogr. Surv. Pure Appl. Math. 64, Longman Scientific & Technical, Harlow, 1993. Copublished in the United States with John Wiley & Sons, New York, NY. (Cited on p. 130.)

[96] M. Diehl, H. G. Bock, and E. Kostina, *An approximation technique for robust nonlinear optimization*, Math. Program., 107 (2006), pp. 213–230, https://doi.org/10.1007/s10107-005-0685-1. (Cited on p. 10.)

[97] J. Dieudonné, *Foundations of Modern Analysis*, Academic Press, New York, NY, 1969. (Cited on pp. 52, 57, and 64.)

[98] M. X. Dong and R. J.-B. Wets, *Estimating density functions: a constrained maximum likelihood approach*, J. Nonparametr. Statist., 12 (2000), pp. 549–595, https://doi.org/10.1080/10485250008832822. (Cited on pp. 72 and 102.)

[99] J. C. Duchi, P. L. Bartlett, and M. J. Wainwright, *Randomized smoothing for stochastic optimization*, SIAM J. Optim., 22 (2012), pp. 674–701, https://doi.org/10.1137/110831659. (Cited on pp. 70, 75, and 108.)

[100] J. C. Duchi and H. Namkoong, *Variance-based Regularization with Convex Objectives*, J. Mach. Learn. Res., 20 (2019), pp. 1–55. (Cited on p. 2.)

[101] L. Dümbgen, S. A. van de Geer, M. C. Veraar, and J. A. Wellner, *Nemirovski's Inequalities Revisited*, Amer. Math. Monthly, 117 (2010), pp. 138–160, https://doi.org/10.4169/000298910X476059. (Cited on pp. 112 and 113.)

[102] I. Dunning, J. Huchette, and M. Lubin, *JuMP: A Modeling Language for Mathematical Optimization*, SIAM Rev., 59 (2017), pp. 295–320, https://doi.org/10.1137/15M1020575. (Cited on p. 38.)

[103] J. Dutta, K. Deb, R. Tulshyan, and R. Arora, *Approximate KKT points and a proximity measure for termination*, J. Global Optim., 56 (2013), pp. 1463–1499, https://doi.org/10.1007/s10898-012-9920-5. (Cited on p. 12.)

[104] M. Eigel, C. Merdon, and J. Neumann, *An adaptive multilevel Monte Carlo method with stochastic bounds for quantities of interest with uncertain data*, SIAM/ASA J. Uncertainty Quantification, 4 (2016), pp. 1219–1245, https://doi.org/10.1137/15M1016448. (Cited on p. 108.)

[105] P. Embrechts, C. Klüppelberg, and T. Mikosch, *Modelling Extremal Events*, Appl. Math. 33, Springer, Berlin, 1997, https://doi.org/10.1007/978-3-642-33483-2. (Cited on p. 123.)

[106] E. Erdoğan and G. Iyengar, *Ambiguous chance constrained problems and robust optimization*, Math. Program., 107 (2006), pp. 37–61, https://doi.org/10.1007/s10107-005-0678-0. (Cited on p. 2.)

[107] P. M. Esfahani and D. Kuhn, *Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations*, Math. Program., 171 (2018), pp. 115–166, https://doi.org/10.1007/s10107-017-1172-1. (Cited on pp. 1, 2, 5, 8, and 9.)

[108] P. E. Farrell, M. B. Giles, M. Croci, T. Roy, and C. Beentjes, *pymlmc: A Python implementation of MLMC: http://people.maths.ox.ac.uk/~gilesm/mlmc/*, Aug. 2015, https://bitbucket.org/pefarrell/pymlmc/. Last updated July 10, 2020, Accessed September 1, 2020. (Cited on p. 123.)

[109] A. V. Fiacco and Y. Ishizuka, *Sensitivity and stability analysis for nonlinear programming*, Ann. Oper. Res., 27 (1990), pp. 215–235, https://doi.org/10.1007/BF02055196. (Cited on p. 24.)

[110] J. Fiala, M. Kočvara, and M. Stingl, *PENLAB: A MATLAB solver for nonlinear semidefinite optimization*, 2013, https://arxiv.org/abs/1311.5240. (Cited on pp. 3, 9, 28, and 29.)

[111] O. E. Flippo and B. Jansen, *Duality and sensitivity in nonconvex quadratic optimization over an ellipsoid*, Eur. J. Oper. Res., 94 (1996), pp. 167–178, https://doi.org/10.1016/0377-2217(95)00199-9. (Cited on p. 9.)

[112] A. Forsgren, P. E. Gill, and M. H. Wright, *Interior methods for nonlinear optimization*, SIAM Rev., 44 (2002), pp. 525–597, https://doi.org/10.1137/S0036144502414942. (Cited on p. 17.)

[113] R. Fortet and E. Mourier, *Les fonctions aléatoires comme éléments aléatoires dans les espaces de Banach*, Studia Math., 15 (1955), pp. 62–79, https://doi.org/10.4064/sm-15-1-62-79. (Cited on p. 113.)

[114] S. Foss, D. Korshunov, and S. Zachary, *An Introduction to Heavy-Tailed and Subexponential Distributions*, Springer Ser. Oper. Res. Financ. Eng., Springer, New York, 2nd ed., 2013, https://doi.org/10.1007/978-1-4614-7101-1. (Cited on p. 123.)

[115] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, Appl. Numer. Harmon. Anal., Birkhäuser, Bosten, 2013, https://doi.org/10.1007/978-0-8176-4948-7. (Cited on p. 5.)

[116] A. L. Fradkov and V. A. Yakubovich, *The S-procedure and duality relations in nonconvex problems of quadratic programming*, Vestnik Leningrad. Univ. Math., 6 (1979), pp. 101–109. In Russian, 1973. (Cited on p. 9.)

[117] D. Kh. Fuk and S. V. Nagaev, *Probability inequalities for sums of independent random variables*, Theory Probab. Appl., 16 (1971), pp. 643–660, https://doi.org/10.1137/1116071. (Cited on p. 129.)

[118] R. Fukuda, *Exponential integrability of sub-Gaussian vectors*, Probab. Theory Related Fields, 85 (1990), pp. 505–521, https://doi.org/10.1007/BF01203168. (Cited on pp. 108, 110, and 111.)

[119] R. Gao and A. J. Kleywegt, *Distributionally robust stochastic optimization with Wasserstein distance*, 2016, https://arxiv.org/abs/1604.02199. (Cited on pp. 1, 2, 9, and 10.)

[120] R. Gao and A. J. Kleywegt, *Distributionally robust stochastic optimization with dependence structure*, 2017, https://arxiv.org/abs/1701.04200. (Cited on p. 1.)

[121] S. Garreis, *Optimal Control under Uncertainty: Theory and Numerical Solution with Low-Rank Tensors*, Dissertation, Technische Universität München, München, 2019, http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:bvb:91-diss-20190215-1452538-1-1. (Cited on pp. 73, 84, and 86.)

[122] S. Garreis, T. M. Surowiec, and M. Ulbrich, *An interior-point approach for solving risk-averse PDE-constrained optimization problems with coherent risk measures*. Preprint No. IGDK-2019-05, Technische Universität München, München, Mar. 2019, http://go.tum.de/498091. (Cited on pp. 42, 83, 86, and 96.)

[123] S. Garreis and M. Ulbrich, *Constrained optimization with low-rank tensors and applications to parametric problems with PDEs*, SIAM J. Sci. Comput., 39 (2017), pp. A25–A54, https://doi.org/10.1137/16M1057607. (Cited on pp. 41, 73, and 101.)

[124] S. Garreis and M. Ulbrich, *A fully adaptive method for the optimal control of semilinear elliptic PDEs under uncertainty using low-rank tensors*, Preprint, Technische Universität München, München, 2019, http://go.tum.de/204409. (Cited on pp. 73, 84, 86, and 101.)

[125] L. Ge, L. Wang, and Y. Chang, *A sparse grid stochastic collocation upwind finite volume element method for the constrained optimal control problem governed by random convection diffusion equations*, J. Sci. Comput., 77 (2018), pp. 524–551, https://doi.org/10.1007/s10915-018-0713-y. (Cited on pp. 4, 70, 85, and 101.)

[126] C. Geiersbach, *Stochastic Approximation for PDE-Constrained Optimization under Uncertainty*, Dissertation, Universität Wien, Wien, 2020. (Cited on p. 74.)

[127] C. Geiersbach, E. Loayza-Romero, and K. Welker, *Stochastic approximation for optimization in shape spaces*, 2020, https://arxiv.org/abs/2001.10786. (Cited on p. 72.)

[128] C. Geiersbach and G. Ch. Pflug, *Projected Stochastic Gradients for Convex Constrained Problems in Hilbert Spaces*, SIAM J. Optim., 29 (2019), pp. 2079–2099, https://doi.org/10.1137/18M1200208. (Cited on pp. 72, 73, 74, 80, 83, and 101.)

[129] C. Geiersbach and T. Scarinci, *Stochastic proximal gradient methods for nonconvex problems in Hilbert spaces*, 2020, https://arxiv.org/abs/2001.01329. (Cited on pp. 72 and 84.)

[130] C. Geiersbach and W. Wollner, *Optimality conditions for convex stochastic optimization problems in Banach spaces with almost sure state constraint*. WIAS Preprint No. 2755, Aug. 2020, https://doi.org/10.20347/WIAS.PREPRINT.2755. (Cited on p. 41.)

[131] C. Geiersbach and W. Wollner, *A stochastic gradient method with mesh refinement for PDE-constrained optimization under uncertainty*, SIAM J. Sci. Comput., 42 (2020), pp. A2750–A2772, https://doi.org/10.1137/19M1263297. (Cited on pp. 70, 73, 74, 80, 83, 86, 100, and 101.)

[132] R. G. Ghanem and P. D. Spanos, *Stochastic Finite Elements: A Spectral Approach*, Springer, New York, NY, 1991, https://doi.org/10.1007/978-1-4612-3094-6. (Cited on p. 1.)

[133] M. B. Giles, *Multilevel Monte Carlo Path Simulation*, Oper. Res., 56 (2008), pp. 607–617, https://doi.org/10.1287/opre.1070.0496. (Cited on pp. 107 and 116.)

[134] M. B. Giles, *Multilevel Monte Carlo methods*, Acta Numer., 24 (2015), pp. 259–328, https://doi.org/10.1017/S096249291500001X. Revised version available at https://people.maths.ox.ac.uk/gilesm/files/acta15.pdf. (Cited on pp. 1, 4, 105, 106, 107, 108, 115, 118, 123, 124, 125, 126, and 129.)

[135] M. B. Giles, T. Nagapetyan, and K. Ritter, *Multilevel Monte Carlo approximation of distribution functions and densities*, SIAM/ASA J. Uncertainty Quantification, 3 (2015), pp. 267–295, https://doi.org/10.1137/140960086. (Cited on p. 113.)

[136] J. Goh and M. Sim, *Distributionally robust optimization and its tractable approximations*, Oper. Res., 58 (2010), pp. 902–917, https://doi.org/10.1287/opre.1090.0795. (Cited on pp. 1 and 8.)

[137] I. E. Grossmann and R. W. H. Sargent, *Optimum design of chemical plants with uncertain parameters*, AIChE J., 24 (1978), pp. 1021–1028, https://doi.org/10.1002/aic.690240612. (Cited on p. 2.)

[138] V. Guigues, A. Juditsky, and A. Nemirovski, *Non-asymptotic confidence bounds for the optimal value of a stochastic program*, Optim. Methods Softw., 32 (2017), pp. 1033–1058, https://doi.org/10.1080/10556788.2017.1350177. (Cited on pp. 70, 71, 75, 77, 81, 82, 108, and 118.)

[139] M. D. Gunzburger, H.-C. Lee, and J. Lee, *Error estimates of stochastic optimal Neumann boundary control problems*, SIAM J. Numer. Anal., 49 (2011), pp. 1532–1552, https://doi.org/10.1137/100801731. (Cited on pp. 70 and 83.)

[140] 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*, 2019, https://arxiv.org/abs/1910.10022. (Cited on p. 73.)

[141] E. Haber, M. Chung, and F. Herrmann, *An effective method for parameter estimation with PDE constraints with multiple right-hand sides*, SIAM J. Optim., 22 (2012), pp. 739–757, https://doi.org/10.1137/11081126X. (Cited on p. 73.)

[142] G. Haeser and M. L. Schuverdt, *On Approximate KKT Condition and its Extension to Continuous Variational Inequalities*, J. Optim. Theory Appl., 149 (2011), pp. 528–539, https://doi.org/10.1007/s10957-011-9802-x. (Cited on p. 12.)

[143] A.-L. Haji-Ali, F. Nobile, E. von Schwerin, and R. Tempone, *Optimization of mesh hierarchies in multilevel Monte Carlo samplers*, Stoch. PDE: Anal. Comp., 4 (2015), pp. 76–112, https://doi.org/10.1007/s40072-015-0049-7. (Cited on p. 108.)

[144] E. T. Hale and Y. Zhang, *Case Studies for a First-Order Robust Nonlinear Programming Formulation*, J. Optim. Theory Appl., 134 (2007), pp. 27–45, https://doi.org/10.1007/s10957-007-9208-y. (Cited on p. 10.)

[145] O. Hanner, *On the uniform convexity of L^p and l^p*, Ark. Mat., 3 (1956), pp. 239–244, https://doi.org/10.1007/BF02589410. (Cited on p. 129.)

[146] S. Heinrich, *Multilevel Monte Carlo Methods*, in Large-Scale Scientific Computing, S. Margenov, J. Waśniewski, and P. Yalamov, eds., Lecture Notes in Comput. Sci. 2179, Berlin, 2001, Springer, pp. 58–67, https://doi.org/10.1007/3-540-45346-6_5. (Cited on pp. 4, 105, 108, and 113.)

[147] L. Hertlein and M. Ulbrich, *An Inexact Bundle Algorithm for Nonconvex Nonsmooth Minimization in Hilbert Space*, SIAM J. Control Optim., 57 (2019), pp. 3137–3165, https://doi.org/10.1137/18M1221849. (Cited on p. 41.)

[148] R. Herzog and F. Schmidt, *Weak lower semi-continuity of the optimal value function and applications to worst-case robust optimal control problems*, Optimization, 61 (2012), pp. 685–697, https://doi.org/10.1080/02331934.2011.603322. (Cited on pp. 41 and 48.)

[149] N. J. Higham and T. Mary, *A New Approach to Probabilistic Rounding Error Analysis*, SIAM J. Sci. Comput., 41 (2019), pp. A2815–A2835, https://doi.org/10.1137/18M1226312. (Cited on p. 5.)

[150] E. Hille and R. S. Phillips, *Functional Analysis and Semi-Groups*, Colloq. Publ. 31, American Mathematical Society, Providence, RI, 1957. (Cited on pp. 46, 53, 85, and 88.)

[151] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich, *Optimization with PDE Constraints*, Math. Model. Theory Appl. 23, Springer, Dordrecht, 2009, https://doi.org/10.1007/978-1-4020-8839-1. (Cited on pp. ix, 44, 45, 48, 49, 51, 55, 57, 58, 59, 63, 64, 65, 70, 72, 77, 78, 83, 84, 85, 86, 89, 90, 91, 94, 101, 120, 121, and 122.)

[152] W. Hoeffding, *Probability Inequalities for Sums of Bounded Random Variables*, J. Amer. Statist. Assoc., 58 (1963), pp. 13–30, https://doi.org/10.2307/2282952. (Cited on p. 81.)

[153] M. Hoffhues, W. Römisch, and T. M. Surowiec, *On quantitative stability in infinite-dimensional optimization under uncertainty*. Preprint No. SPP1962-131, WIAS, Berlin, Jan. 2020, https://spp1962.wias-berlin.de/preprints/131.pdf. (Cited on p. 72.)

[154] W. W. Hogan, *Point-to-set maps in mathematical programming*, SIAM Rev., 15 (1973), pp. 591–603, https://doi.org/10.1137/1015073. (Cited on pp. 21, 43, and 44.)

[155] R. B. Holmes, *Geometric Functional Analysis and its Applications*, Grad. Texts in Math. 24, Springer, New York, NY, 1975, https://doi.org/10.1007/978-1-4684-9369-6. (Cited on pp. 130 and 131.)

[156] T. Homem-de Mello and G. Bayraksan, *Stochastic Constraints and Variance Reduction Techniques*, in Handbook of Simulation Optimization, M. C. Fu, ed., Springer, New York, NY, 2015, pp. 245–276, https://doi.org/10.1007/978-1-4939-1384-8_9. (Cited on p. 69.)

[157] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 2nd ed., 2013. (Cited on pp. 13, 14, 17, 22, 43, and 48.)

[158] B. Houska and M. Diehl, *Nonlinear Robust Optimization via Sequential Convex Bilevel Programming*, Math. Program., 142 (2013), pp. 539–577, https://doi.org/10.1007/s10107-012-0591-2. (Cited on pp. 10, 11, and 33.)

[159] T. Hytönen, J. van Neerven, M. Veraar, and L. Weis, *Analysis in Banach Spaces: Martingales and Littlewood-Paley Theory*, Ergeb. Math. Grenzgeb. (3) 63, Springer, Cham, 2016, https://doi.org/10.1007/978-3-319-48520-1. (Cited on pp. ix, 46, 57, 84, 86, 87, 88, 90, 93, 109, 110, 112, 122, and 132.)

[160] T. Hytönen, J. van Neerven, M. Veraar, and L. Weis, *Analysis in Banach Spaces: Probabilistic Methods and Operator Theory*, Ergeb. Math. Grenzgeb. (3) 67, Springer, Cham, 2017, https://doi.org/10.1007/978-3-319-69808-3. (Cited on pp. 88 and 122.)

[161] A. D. Ioffe and V. L. Levin, *Subdifferentials of convex functions*, Tr. Mosk. Mat. Obs., 26 (1972), pp. 3–73, http://mi.mathnet.ru/eng/mmo257. (Cited on pp. 96, 98, and 99.)

[162] I. G. Ion, Z. Bontinck, D. Loukrezis, U. Römer, O. Lass, S. Ulbrich, S. Schöps, and H. De Gersem, *Robust shape optimization of electric devices based on deterministic optimization methods and finite-element analysis with affine parametrization and design elements*, Electr. Eng., 100 (2018), pp. 2635–2647, https://doi.org/10.1007/s00202-018-0716-6. (Cited on p. 27.)

[163] K. Ito and K. Kunisch, *Augmented Lagrangian methods for nonsmooth, convex optimization in Hilbert spaces*, Nonlinear Anal., 41 (2000), pp. 591–616, https://doi.org/10.1016/S0362-546X(98)00299-5. (Cited on pp. 78, 79, 84, 93, and 98.)

[164] M. Ivanov and S. Troyanski, *Uniformly smooth renorming of Banach spaces with modulus of convexity of power type 2*, J. Funct. Anal., 237 (2006), pp. 373–390, https://doi.org/10.1016/j.jfa.2006.03.024. (Cited on p. 132.)

[165] B. Jourdain and A. Kebaier, *Non-asymptotic error bounds for the multilevel Monte Carlo Euler method applied to SDEs with constant diffusion coefficient*, Electron. J. Probab., 24 (2019), pp. 1–34, https://doi.org/10.1214/19-EJP271. (Cited on p. 108.)

[166] A. Juditsky and A. Nemirovski, *Large deviations of vector-valued martingales in 2-smooth normed spaces*, 2008, https://arxiv.org/abs/0809.0813. (Cited on pp. 73, 105, 111, 112, 113, 115, 129, 130, 132, and 134.)

[167] A. Juditsky and A. Nemirovski, *Statistical Inference via Convex Optimization*, Princeton Ser. Appl. Math., Princeton University Press, Princeton, NJ, 2020, https://doi.org/10.2307/j.ctvqsdxqd. (Cited on p. 5.)

[168] A. Juditsky and Yu. Nesterov, *Deterministic and stochastic primal-dual subgradient algorithms for uniformly convex minimization*, Stoch. Syst., 4 (2014), pp. 44–80, https://doi.org/10.1214/10-SSY010. (Cited on p. 72.)

[169] O. Kallenberg, *Foundations of Modern Probability*, Probab. Appl., Springer, New York, NY, 2nd ed., 2002, https://doi.org/10.1007/978-1-4757-4015-8. (Cited on pp. 44, 45, 46, 47, 52, 78, 85, 86, and 102.)

[170] Yu. M. Kaniovski, A. J. King, and R. J.-B. Wets, *Probabilistic bounds (via large deviations) for the solutions of stochastic programming problems*, Ann. Oper. Res., 56 (1995), pp. 189–208, https://doi.org/10.1007/BF02031707. (Cited on pp. 72 and 102.)

[171] C. Kanzow and A. Schwartz, *The price of inexactness: Convergence properties of relaxation methods for mathematical programs with complementarity constraints revisited*, Math. Oper. Res., 40 (2014), pp. 253–275, https://doi.org/10.1287/moor.2014.0667. (Cited on p. 10.)

[172] F. Kappel and A. V. Kuntsevich, *SolvOpt: The solver for local nonlinear optimization problems*. Institute for Mathematics, University of Graz, Graz, 1997, https://imsc.uni-graz.at/kuntsevich/solvopt/index.html. Accessed September 12, 2019. (Cited on p. 34.)

[173] F. Kappel and A. V. Kuntsevich, *An implementation of Shor's r-algorithm*, Comput. Optim. Appl., 15 (2000), pp. 193–205, https://doi.org/10.1023/A:1008739111712. (Cited on p. 34.)

[174] N. Karmitsa, *Solver-o-matic*. http://napsu.karmitsa.fi/solveromatic/, 2012. Accessed September 12, 2019. (Cited on p. 28.)

[175] N. Karmitsa, A. Bagirov, and M. M. Mäkelä, *Comparing different nonsmooth minimization methods and software*, Optim. Methods Softw., 27 (2012), pp. 131–153, https://doi.org/10.1080/10556788.2010.526116. (Cited on p. 9.)

[176] A. Kebaier and J. Lelong, *Coupling Importance Sampling and Multilevel Monte Carlo using Sample Average Approximation*, Methodol. Comput. Appl. Probab., 20 (2018), pp. 611–641, https://doi.org/10.1007/s11009-017-9579-y. (Cited on p. 108.)

[177] S. Kim, R. Pasupathy, and S. G. Henderson, *A Guide to Sample Average Approximation*, in Handbook of Simulation Optimization, M. C. Fu, ed., Springer, New York, NY, 2015, pp. 207–243, https://doi.org/10.1007/978-1-4939-1384-8_8. (Cited on p. 69.)

[178] D. Kinderlehrer and G. Stampacchia, *An Introduction to Variational Inequalities and Their Applications*, Classics Appl. Math. 31, SIAM, Philadelphia, PA, 2000, https://doi.org/10.1137/1.9780898719451. (Cited on p. 93.)

[179] A. J. Kleywegt, A. Shapiro, and T. Homem-de Mello, *The sample average approximation method for stochastic discrete optimization*, SIAM J. Optim., 12 (2002), pp. 479–502, https://doi.org/10.1137/S1052623499363220. (Cited on p. 69.)

[180] P. Kolvenbach, *Robust optimization of PDE-constrained problems using second-order models and nonsmooth approaches*, Dissertation, Technische Universität Darmstadt, Darmstadt, 2018. (Cited on pp. 10, 33, and 66.)

[181] P. Kolvenbach, O. Lass, and S. Ulbrich, *An approach for robust PDE-constrained optimization with application to shape optimization of electrical engines and of dynamic elastic structures under uncertainty*, Optim. Eng., 19 (2018), pp. 697–731, https://doi.org/10.1007/s11081-018-9388-3. (Cited on pp. 10, 33, 41, 42, and 66.)

[182] R. Kornhuber, Ch. Schwab, and M.-W. Wolf, *Multilevel Monte Carlo Finite Element Methods for Stochastic Elliptic Variational Inequalities*, SIAM J. Numer. Anal., 52 (2014), pp. 1243–1268, https://doi.org/10.1137/130916126. (Cited on p. 108.)

[183] R. Kornhuber and E. Youett, *Adaptive Multilevel Monte Carlo Methods for Stochastic Variational Inequalities*, SIAM J. Numer. Anal., 56 (2018), pp. 1987–2007, https://doi.org/10.1137/16M1104986. (Cited on p. 108.)

[184] D. P. Kouri, *An Approach for the Adaptive Solution of Optimization Problems Governed by Partial Differential Equations with Uncertain Coefficients*, PhD thesis, Rice University, Houston, TX, 2012, https://apps.dtic.mil/dtic/tr/fulltext/u2/a577917.pdf. (Cited on pp. 70, 73, 83, and 100.)

[185] D. P. Kouri, *A Multilevel Stochastic Collocation Algorithm for Optimization of PDEs with Uncertain Coefficients*, SIAM/ASA J. Uncertainty Quantification, 2 (2014), pp. 55–81, https://doi.org/10.1137/130915960. (Cited on pp. 51 and 73.)

[186] D. P. Kouri, *A Measure Approximation for Distributionally Robust PDE-Constrained Optimization Problems*, SIAM J. Numer. Anal., 55 (2017), pp. 3147–3172, https://doi.org/10.1137/15M1036944. (Cited on pp. 1 and 41.)

[187] D. P. Kouri, *Spectral risk measures: the risk quadrangle and optimal approximation*, Math. Program., 174 (2019), pp. 525–552, https://doi.org/10.1007/s10107-018-1267-3. (Cited on pp. 73 and 96.)

[188] D. P. Kouri, M. Heinkenschloss, D. Ridzal, and B. van Bloemen Waanders, *A Trust-Region Algorithm with Adaptive Stochastic Collocation for PDE Optimization under Uncertainty*, SIAM J. Sci. Comput., 35 (2013), pp. A1847–A1879, https://doi.org/10.1137/120892362. (Cited on pp. 51, 52, 53, 54, and 73.)

[189] D. P. Kouri, M. Heinkenschloss, D. Ridzal, and B. G. van Bloemen Waanders, *Inexact objective function evaluations in a trust-region algorithm for PDE-constrained optimization under uncertainty*, SIAM J. Sci. Comput., 36 (2014), pp. A3011–A3029, https://doi.org/10.1137/140955665. (Cited on pp. 41, 73, and 101.)

[190] D. P. Kouri and A. Shapiro, *Optimization of PDEs with Uncertain Inputs*, in Frontiers in PDE-Constrained Optimization, H. Antil, D. P. Kouri, M.-D. Lacasse, and D. Ridzal, eds., IMA Vol. Math. Appl. 163, Springer, New York, NY, 2018, pp. 41–81, https://doi.org/10.1007/978-1-4939-8636-1_2. (Cited on pp. 1, 41, 44, 69, 73, 74, 80, 83, 85, and 96.)

[191] 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. (Cited on pp. 41, 42, 51, 54, 55, 83, 96, and 100.)

[192] D. P. Kouri and T. M. Surowiec, *Existence and Optimality Conditions for Risk-Averse PDE-Constrained Optimization*, SIAM/ASA J. Uncertainty Quantification, 6 (2018), pp. 787–815, https://doi.org/10.1137/16M1086613. (Cited on pp. 41, 44, 74, 85, and 96.)

[193] D. P. Kouri and T. M. Surowiec, *Epi-Regularization of Risk Measures*, Math. Oper. Res., 45 (2020), pp. 774–795, https://doi.org/10.1287/moor.2019.1013. (Cited on pp. 42, 62, and 96.)

[194] D. P. Kouri and T. M. Surowiec, *A primal-dual algorithm for risk minimization*, Nov. 2020, https://doi.org/10.13140/RG.2.2.31187.20004. (Cited on pp. 72 and 96.)

[195] D. P. Kouri and T. M. Surowiec, *Risk-averse optimal control of semilinear elliptic PDEs*, ESAIM Control. Optim. Calc. Var., 26 (2020), https://doi.org/10.1051/cocv/2019061. (Cited on p. 96.)

[196] E. Kreyszig, *Introductory Functional Analysis with Applications*, John Wiley & Sons, New York, NY, 1978. (Cited on pp. 51, 52, 54, 63, 72, 84, 86, 88, and 90.)

[197] S. Krumscheid and F. Nobile, *Multilevel Monte Carlo Approximation of Functions*, SIAM/ASA J. Uncertainty Quantification, 6 (2018), pp. 1256–1293, https://doi.org/10.1137/17M1135566. (Cited on p. 113.)

[198] F. Kruse and M. Ulbrich, *A self-concordant interior point approach for optimal control with state constraints*, SIAM J. Optim., 25 (2015), pp. 770–806, https://doi.org/10.1137/130936671. (Cited on p. 42.)

[199] M. Kuchlbauer, F. Liers, and M. Stingl, *An adaptive bundle method for nonlinear robust optimization.* Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, 2020, https://opus4.kobv.de/opus4-trr154/files/307/manuscript_kuchlbauer.pdf. (Cited on pp. 10 and 33.)

[200] J. Kuelbs, *An inequality for the distribution of a sum of certain Banach space valued random variables*, Studia Math., 52 (1974), pp. 69–87, https://doi.org/10.4064/sm-52-1-69-87. (Cited on p. 112.)

[201] J. Kuelbs and T. Kurtz, *Berry–Esséen estimates in Hilbert space and an application to the law of the iterated logarithm*, Ann. Probab., 2 (1974), pp. 387–407, https://doi.org/10.1214/aop/1176996655. (Cited on p. 109.)

[202] K. Kunisch and D. Walter, *Semiglobal optimal feedback stabilization of autonomous systems via deep neural network approximation*, 2020, https://arxiv.org/abs/2002.08625. (Cited on p. 63.)

[203] F. Y. Kuo and D. Nuyens, *Application of quasi-Monte Carlo methods to elliptic PDEs with random diffusion coefficients: A survey of analysis and implementation*, Found. Comput. Math., 16 (2016), pp. 1631–1696, https://doi.org/10.1007/s10208-016-9329-5. (Cited on p. 108.)

[204] F. Y. Kuo, Ch. Schwab, and I. H. Sloan, *Multi-level quasi-Monte Carlo finite element methods for a class of elliptic PDEs with random coefficients*, Found. Comput. Math., 15 (2015), pp. 411–449, https://doi.org/10.1007/s10208-014-9237-5. (Cited on p. 108.)

[205] V. Kvaratskhelia, V. Tarieladze, and N. Vakhania, *Characterization of γ-Subgaussian Random Elements in a Banach Space*, J. Math. Sci., 216 (2016), pp. 564–568, https://doi.org/10.1007/s10958-016-2915-x. (Cited on p. 110.)

[206] B. M. Kwak and E. J. Haug, *Optimum design in the presence of parametric uncertainty*, J. Optim. Theory Appl., 19 (1976), pp. 527–546, https://doi.org/10.1007/BF00934653. (Cited on p. 2.)

[207] G. Lan, *First-order and Stochastic Optimization Methods for Machine Learning*, Springer Ser. Data Sci., Springer, Cham, 2020, https://doi.org/10.1007/978-3-030-39568-1. (Cited on pp. 70, 72, and 75.)

[208] G. Lan, A. Nemirovski, and A. Shapiro, *Validation analysis of mirror descent stochastic approximation method*, Math. Program., 134 (2012), pp. 425–458, https://doi.org/10.1007/s10107-011-0442-6. (Cited on pp. 5, 72, and 97.)

[209] 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. (Cited on pp. 10, 27, 33, 41, and 42.)

[210] R. Latała and K. Oleszkiewicz, *Gaussian measures of dilatations of convex symmetric sets*, Ann. Probab., 27 (1999), pp. 1922–1938. (Cited on p. 135.)

[211] P. D. Lax, *Functional Analysis*, Pure Appl. Math., Wiley-Interscience, New York, NY, 2002. (Cited on pp. viii and 111.)

[212] M. Ledoux, *Sur les théorèmes limites dans certains espaces de Banach lisses*, in Probability in Banach spaces IV, Lecture Notes in Math. 990, Springer, Berlin, 1983, pp. 150–169, https://doi.org/10.1007/BFb0064269. Proceedings of the Seminar Held in Oberwolfach, Germany, July 1982. (Cited on pp. 108, 111, 113, and 132.)

[213] V. L. Levin, *Convex integral functionals and the theory of lifting*, Russ. Math. Surv., 30 (1975), pp. 119–184, https://doi.org/10.1070/rm1975v030n02abeh001408. (Cited on p. 102.)

[214] A. S. Lewis, *Derivatives of Spectral Functions*, Math. Oper. Res., 21 (1996), pp. 576–588, https://doi.org/10.1287/moor.21.3.576. (Cited on p. 14.)

[215] A. S. Lewis, *Nonsmooth analysis of eigenvalues*, Math. Program., 84 (1999), pp. 1–24, https://doi.org/10.1007/s10107980004a. (Cited on pp. 9 and 15.)

[216] A. S. Lewis and M. L. Overton, *Nonsmooth optimization via quasi-Newton methods*, Math. Program., 141 (2013), pp. 135–163, https://doi.org/10.1007/s10107-012-0514-2. (Cited on p. 9.)

[217] A. S. Lewis and H. S. Sendov, *Twice differentiable spectral functions*, SIAM J. Matrix Anal. Appl., 23 (2001), pp. 368–386, https://doi.org/10.1137/S089547980036838X. (Cited on pp. 13, 17, 43, and 50.)

[218] S. Leyffer, M. Menickelly, T. Munson, C. Vanaret, and S. M. Wild, *A survey of nonlinear robust optimization*, INFOR Inf. Syst. Oper. Res., 58 (2020), pp. 342–373, https://doi.org/10.1080/03155986.2020.1730676. (Cited on pp. 10 and 33.)

[219] M. Liu, R. Kumar, E. Haber, and A. Aravkin, *Simultaneous-shot inversion for PDE-constrained optimization problems with missing data*, Inverse Probl., 35 (2018), p. 025003, https://doi.org/10.1088/1361-6420/aaf317. (Cited on p. 73.)

[220] A. Logg, K.-A. Mardal, and G. N. Wells, eds., *Automated Solution of Differential Equations by the Finite Element Method*, Lect. Notes Comput. Sci. Eng. 84, Springer, Heidelberg, 2012, https://doi.org/10.1007/978-3-642-23099-8. (Cited on pp. 40, 54, and 124.)

[221] G. J. Lord, C. E. Powell, and T. Shardlow, *An Introduction to Computational Stochastic PDEs*, Cambridge Texts Appl. Math. 50, Cambridge University Press, Cambridge, 2014, https://doi.org/10.1017/CBO9781139017329. (Cited on pp. 109 and 122.)

[222] G. Lugosi and S. Mendelson, *Mean Estimation and Regression Under Heavy-Tailed Distributions: A Survey*, Found. Comput. Math., 19 (2019), pp. 1145–1190, https://doi.org/10.1007/s10208-019-09427-x. (Cited on p. 129.)

[223] M. M. Mäkelä, *Multiobjective proximal bundle method for nonconvex nonsmooth optimization: Fortran subroutine MPBNGC 2.0*, Reports of the Department of Mathematical Information Technology, Series B. Scientific Computing B 13/2003, University of Jyväskylä, Jyväskylä, 2003. (Cited on pp. 3, 8, 9, and 28.)

[224] M. M. Mäkelä, N. Karmitsa, and O. Wilppu, *Proximal Bundle Method for Nonsmooth and Nonconvex Multiobjective Optimization*, in Mathematical Modeling and Optimization of Complex Structures, P. Neittaanmäki, S. Repin, and T. Tuovinen, eds., Comput. Methods Appl. Sci. 40, Springer, Cham, 2016, pp. 191–204, https://doi.org/10.1007/978-3-319-23564-6_12. (Cited on pp. 3, 8, 9, 28, and 29.)

[225] M. M. Mäkelä and P. Neittaanmäki, *Nonsmooth Optimization: Analysis and Algorithms with Applications to Optimal Control*, World Scientific, Singapore, 1992, https://doi.org/10.1142/1493. (Cited on pp. 11 and 24.)

[226] F. J. Marín, J. Martínez-Frutos, and F. Periago, *Control of Random PDEs: An Overview*, in Recent Advances in PDEs: Analysis, Numerics and Control: In Honor of Prof. Fernández-Cara's 60th Birthday, A. Doubova, M. González-Burgos, F. Guillén-González, and M. Marín Beltrán, eds., Springer, Cham, 2018, pp. 193–210, https://doi.org/10.1007/978-3-319-97613-6_10. (Cited on pp. 4 and 86.)

[227] M. Martin, S. Krumscheid, and F. Nobile, *Analysis of stochastic gradient methods for PDE-constrained optimal control problems with uncertain parameters*, tech. report, École Polytechnique Fédérale de Lausanne, Lausanne, 2018, https://doi.org/10.5075/epfl-MATHICSE-263568. MATHICSE Technical Report Nr. 04.2018, Mar. 2018. (Cited on pp. 70, 73, 83, and 86.)

[228] M. C. Martin, *Stochastic approximation methods for PDE constrained optimal control problems with uncertain parameters*, PhD thesis, École Polytechnique Fédérale de Lausanne, Lausanne, 2019, https://doi.org/10.5075/epfl-thesis-7233. (Cited on pp. 73 and 83.)

[229] M. C. Martin, F. Nobile, and P. Tsilifis, *A multilevel stochastic gradient method for PDE-constrained optimal control problems with uncertain parameters*, tech. report, École Polytechnique Fédérale de Lausanne, Lausanne, 2020, https://doi.org/10.5075/epfl-MATHICSE-273651. (Cited on p. 73.)

[230] J. Martínez-Frutos and F. P. Esparza, *Optimal Control of PDEs under Uncertainty: An Introduction with Application to Optimal Shape Design of Structures*, SpringerBriefs Math., Springer, Cham, 2018, https://doi.org/10.1007/978-3-319-98210-6. (Cited on pp. 4, 41, 70, 83, and 86.)

[231] D. Meidner and B. Vexler, *A Priori Error Estimates for Space-Time Finite Element Discretization of Parabolic Optimal Control Problems Part II: Problems with Control Constraints*, SIAM J. Control Optim., 47 (2008), pp. 1301–1329, https://doi.org/10.1137/070694028. (Cited on p. 94.)

[232] M. Menickelly and S. M. Wild, *Derivative-free robust optimization by outer approximations*, Math. Program., 179 (2018), pp. 157–193, https://doi.org/10.1007/s10107-018-1326-9. (Cited on p. 10.)

[233] V. Milman, *Geometric theory of Banach spaces, Part II; Geometry of the unit sphere*, Russ. Math. Surv., 26 (1971), pp. 79–163, https://doi.org/10.1070/rm1971v026n06abeh001273. (Cited on pp. 108 and 111.)

[234] J. Milz and M. Ulbrich, *An approximation scheme for distributionally robust nonlinear optimization*, SIAM J. Optim., 30 (2020), pp. 1996–2025, https://doi.org/10.1137/19M1263121. (Cited on pp. v, 3, 7, 12, 13, 16, 17, 18, 20, 21, 22, and 26.)

[235] J. Milz and M. Ulbrich, *An approximation scheme for distributionally robust PDE-constrained optimization.* Preprint No. IGDK-2020-09, Technische Universität München, München, Jun. 2020, http://www.igdk.eu/foswiki/pub/IGDK1754/Preprints/MilzUlbrich-PDEDRO.pdf. (Cited on pp. v, 4, 7, 35, 39, 44, 45, 46, 47, 48, 49, 50, 62, and 63.)

[236] S. Minsker, *Geometric median and robust estimation in Banach spaces*, Bernoulli, 21 (2015), pp. 2308–2335, https://doi.org/10.3150/14-BEJ645. (Cited on p. 129.)

[237] J. J. Moré, B. S. Garbow, and K. E. Hillstrom, *Testing unconstrained optimization software*, ACM Trans. Math. Softw., 7 (1981), pp. 17–41, https://doi.org/10.1145/355934.355936. (Cited on pp. 27, 29, 31, and 34.)

[238] J. J. Moré and D. C. Sorensen, *Computing a Trust Region Step*, SIAM J. Sci. and Stat. Comput., 4 (1983), pp. 553–572, https://doi.org/10.1137/0904038. (Cited on pp. 15, 16, 17, 29, 30, 42, and 54.)

[239] A. Mutapcic and S. Boyd, *Cutting-set methods for robust convex optimization with pessimizing oracles*, Optim. Methods Softw., 24 (2009), pp. 381–406, https://doi.org/10.1080/10556780802712889. (Cited on p. 10.)

[240] A. Nemirovski, *Topics in Non-Parametric Statistics*, in Lectures on Probability Theory and Statistics: Ecole d'Eté de Probabilités de Saint-Flour XXVIII - 1998, P. Bernard, ed., Lecture Notes in Math. 1738, Springer, Berlin, 1998, pp. 88–277, https://doi.org/10.1007/BFb0106703. (Cited on p. 113.)

[241] A. Nemirovski, *On tractable approximations of randomly perturbed convex constraints*, in 42nd IEEE International Conference on Decision and Control, vol. 3, Maui, HI, 2003, pp. 2419–2422, https://doi.org/10.1109/CDC.2003.1272982. (Cited on p. 73.)

[242] A. Nemirovski, *Regular Banach spaces and large deviations of random sums*, tech. report, Faculty of Industrial Engineering and Management, Israel Institute of Technology, Haifa, 2004, https://www2.isye.gatech.edu/~nemirovs/LargeDev2004.pdf. (Cited on pp. 111, 112, and 113.)

[243] A. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro, *Robust Stochastic Approximation Approach to Stochastic Programming*, SIAM J. Optim., 19 (2009), pp. 1574–1609, https://doi.org/10.1137/070704277. (Cited on pp. 5, 70, 72, 75, 80, 83, 101, 108, and 110.)

[244] A. Nemirovski and A. Shapiro, *Scenario Approximations of Chance Constraints*, in Probabilistic and Randomized Methods for Design under Uncertainty, G. Calafiore and F. Dabbene, eds., Springer, London, 2006, pp. 3–47, https://doi.org/10.1007/1-84628-095-8_1. (Cited on p. 5.)

[245] A. Nemirovski and A. Shapiro, *Convex Approximations of Chance Constrained Programs*, SIAM J. Optim., 17 (2007), pp. 969–996, https://doi.org/10.1137/050622328. (Cited on pp. 2 and 129.)

[246] A. S. Nemirovsky and D. B. Yudin, *Problem Complexity and Method Efficiency in Optimization*, Wiley-Interscience Series in Discrete Mathematics, John Wiley & Sons, Chichester, 1983. Translated by E. R. Dawson. (Cited on pp. 72, 75, 78, 83, 86, 101, 111, and 129.)

[247] Yu. Nesterov, *Smoothing technique and its applications in semidefinite optimization*, Math. Program., 110 (2007), pp. 245–259, https://doi.org/10.1007/s10107-006-0001-8. (Cited on pp. 17, 22, and 51.)

[248] Yu. Nesterov, *Primal-dual subgradient methods for convex problems*, Math. Program., 120 (2009), pp. 221–259, https://doi.org/10.1007/s10107-007-0149-x. (Cited on p. 72.)

[249] Yu. Nesterov and A. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming*, SIAM Stud. Appl. Math. 13, SIAM, Philadelphia, PA, 1994, https://doi.org/10.1137/1.9781611970791. (Cited on p. 19.)

[250] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer Ser. Oper. Res. Financ. Eng., Springer, New York, 2nd ed., 2006, https://doi.org/10.1007/978-0-387-40065-5_11. (Cited on p. 33.)

[251] B. O'Donoghue, E. Chu, N. Parikh, and S. Boyd, *Conic Optimization via Operator Splitting and Homogeneous Self-Dual Embedding*, J. Optim. Theory Appl., 169 (2016), pp. 1042–1068, https://doi.org/10.1007/s10957-016-0892-3. (Cited on p. 38.)

[252] B. O'Donoghue, E. Chu, N. Parikh, and S. Boyd, *SCS: Splitting Conic Solver, version 2.1.2*. https://github.com/cvxgrp/scs, Nov. 2019. (Cited on p. 38.)

[253] G. Ch. Pflug and D. Wozabal, *Ambiguity in portfolio selection*, Quant. Financ., 7 (2007), pp. 435–442, https://doi.org/10.1080/14697680701455410. (Cited on pp. 1 and 8.)

[254] C. Phelps, J. 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. (Cited on p. 72.)

[255] I. Pinelis, *An approach to inequalities for the distributions of infinite-dimensional martingales*, in Probability in Banach Spaces, 8: Proceedings of the Eighth International Conference, R. M. Dudley, M. G. Hahn, and J. Kuelbs, eds., Progr. Probab. 30, Birkhäuser, Boston, MA, 1992, pp. 128–134, https://doi.org/10.1007/978-1-4612-0367-4_9. (Cited on p. 109.)

[256] I. Pinelis, *Optimum bounds for the distributions of martingales in Banach spaces*, Ann. Probab., 22 (1994), pp. 1679–1706, https://doi.org/10.1214/aop/1176988477. (Cited on pp. 73, 103, 105, 109, 111, 112, 113, 132, and 134.)

[257] I. Pinelis, *Sharp Exponential Inequalities for the Martingales in the 2-smooth Banach spaces and Applications to "Scalarizing" Decoupling*, in Probability in Banach Spaces, 9, J. Hoffmann-Jørgensen, J. Kuelbs, and M. B. Marcus, eds., Progr. Probab. 35, Birkhäuser, Boston, MA, 1994, pp. 55–70, https://doi.org/10.1007/978-1-4612-0253-0_4. (Cited on pp. 73, 105, 111, and 134.)

[258] I. F. Pinelis, *Inequalities for distributions of sums of independent random vectors and their application to estimating a density*, Theory Probab. Appl., 35 (1991), pp. 605–607, https://doi.org/10.1137/1135088. (Cited on pp. 105, 109, and 133.)

[259] I. F. Pinelis and A. I. Sakhanenko, *Remarks on Inequalities for Large Deviation Probabilities*, Theory Probab. Appl., 30 (1986), pp. 143–148, https://doi.org/10.1137/1130013. (Cited on pp. 71, 73, 103, 105, 109, 111, and 133.)

[260] G. Pisier, *Martingales with values in uniformly convex spaces*, Israel J. Math., 20 (1975), pp. 326–350, https://doi.org/10.1007/BF02760337. (Cited on p. 113.)

[261] G. Pisier, *Probabilistic methods in the geometry of Banach spaces*, in Probability and Analysis, G. Letta and M. Pratelli, eds., Lecture Notes in Math. 1206, Berlin, 1986, Springer, pp. 167–241, https://doi.org/10.1007/BFb0076302. (Cited on p. 109.)

[262] G. Pisier, *Martingales in Banach Spaces*, Cambridge Stud. Adv. Math., Cambridge University Press, Cambridge, 2016, https://doi.org/10.1017/CBO9781316480588. (Cited on pp. 109, 111, 112, 113, and 130.)

[263] E. Polak, J. O. Royset, and R. Womersley, *Algorithms with Adaptive Smoothing for Finite Minimax Problems*, J. Optim. Theory Appl., 119 (2003), pp. 459–484, https://doi.org/10.1023/B:JOTA.0000006685.60019.3e. (Cited on p. 13.)

[264] I. Popescu, *Robust Mean-Covariance Solutions for Stochastic Optimization*, Oper. Res., 55 (2007), pp. 98–112, https://doi.org/10.1287/opre.1060.0353. (Cited on pp. 1, 8, and 9.)

[265] F. Pörner, *Regularization Methods for Ill-Posed Optimal Control Problems*, Dissertation, Universität Würzburg, Würzburg, 2018, https://doi.org/10.25972/WUP-978-3-95826-087-0. (Cited on p. 93.)

[266] J. Revels, M. Lubin, and T. Papamarkou, *Forward-Mode Automatic Differentiation in Julia*, 2016, https://arxiv.org/abs/1607.07892. (Cited on p. 29.)

[267] R. T. Rockafellar, *Convex integral functionals and duality*, in Contributions to Nonlinear Functional Analysis, E. H. Zarantonello, ed., Academic Press, 1971, pp. 215–236, https://doi.org/10.1016/B978-0-12-775850-3.50012-1. (Cited on pp. 74 and 102.)

[268] R. T. Rockafellar, *Integral functionals, normal integrands and measurable selections*, in Nonlinear Operators and the Calculus of Variations, J. P. Gossez, E. J. Lami Dozo, J. Mawhin, and L. Waelbroeck, eds., Lecture Notes in Math. 543, Berlin, 1976, Springer, pp. 157–207, https://doi.org/10.1007/BFb0079944. (Cited on pp. 74 and 102.)

[269] R. T. Rockafellar, *Coherent Approaches to Risk in Optimization Under Uncertainty*, in OR Tools and Applications: Glimpses of Future Technologies, P. Gray, ed., INFORMS, Maryland, 2007, ch. 3, pp. 38–61, https://doi.org/10.1287/educ.1073.0032. (Cited on pp. 1, 2, 41, and 96.)

[270] R. T. Rockafellar, *Solving stochastic programming problems with risk measures by progressive hedging*, Set-Valued Var. Anal., 26 (2018), pp. 759–768, https://doi.org/10.1007/s11228-017-0437-4. (Cited on p. 72.)

[271] R. T. Rockafellar and J. Royset, *On buffered failure probability in design and optimization of structures*, Reliab. Eng. Syst. Safe., 95 (2010), pp. 499–510, https://doi.org/10.1016/j.ress.2010.01.001. (Cited on pp. 10 and 96.)

[272] R. T. Rockafellar and S. Uryasev, *Optimization of Conditional Value-at-Risk*, Journal of Risk, 2 (2000), pp. 21–41, https://doi.org/10.21314/JOR.2000.038. (Cited on p. 96.)

[273] R. T. Rockafellar and S. Uryasev, *Conditional value-at-risk for general loss distributions*, J. Banking Finance, 26 (2002), pp. 1443–1471, https://doi.org/10.1016/S0378-4266(02)00271-6. (Cited on pp. 96, 97, 98, and 100.)

[274] R. T. Rockafellar, S. Uryasev, and M. Zabarankin, *Generalized deviations in risk analysis*, Finance Stoch., 10 (2006), pp. 51–74, https://doi.org/10.1007/s00780-005-0165-8. (Cited on p. 2.)

[275] R. T. Rockafellar and R. J.-B. Wets, *Scenarios and policy aggregation in optimization under uncertainty*, Math. Oper. Res., 16 (1991), pp. 119–147, https://doi.org/10.1287/moor.16.1.119. (Cited on p. 72.)

[276] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Grundlehren Math. Wiss. 317, Springer, Berlin, 2009, https://doi.org/10.1007/978-3-642-02431-3. (Cited on pp. 12 and 102.)

[277] H. L. Royden, *Real Analysis*, Macmillan Publishing Company, New York, NY, 3rd ed., 1988. (Cited on p. 63.)

[278] J. O. Royset, *Approximations of semicontinuous functions with applications to stochastic optimization and statistical estimation*, Math. Program., (2019), https://doi.org/10.1007/s10107-019-01413-z. (Cited on p. 71.)

[279] A. Ruszczyński and A. Shapiro, *Optimization of Convex Risk Functions*, Math. Oper. Res., 31 (2006), pp. 433–452, https://doi.org/10.1287/moor.1050.0186. (Cited on pp. 96 and 100.)

[280] A. Ruszczyński and A. Shapiro, *Optimization of Risk Measures*, in Probabilistic and Randomized Methods for Design under Uncertainty, G. Calafiore and F. Dabbene, eds., Springer, London, 2006, pp. 119–157, https://doi.org/10.1007/1-84628-095-8_4. (Cited on p. 41.)

[281] A. Ruszczyński and A. Shapiro, *Corrigendum to: "Optimization of Convex Risk Functions," Mathematics of Operations Research 31 (2006) 433–452*, Math. Oper. Res., 32 (2007), pp. 496–496, https://doi.org/10.1287/moor.1070.0265. (Cited on p. 74.)

[282] L. L. Sakalauskas, *Nonlinear stochastic programming by Monte-Carlo estimators*, Eur. J. Oper. Res., 137 (2002), pp. 558–573, https://doi.org/10.1016/S0377-2217(01)00109-6. (Cited on p. 72.)

[283] H. Scarf, *A min-max solution of an inventory problem*, in Studies in the Mathematical Theory of Inventory and Production, Stanford University Press, Stanford, 1958, pp. 201–209. (Cited on pp. 1 and 9.)

[284] R. L. Schilling, *Measures, Integrals and Martingales*, Cambridge University Press, Cambridge, 2005. (Cited on p. 102.)

[285] C. Schillings and V. Schulz, *On the influence of robustness measures on shape optimization with stochastic uncertainties*, Optim. Eng., 16 (2015), pp. 347–386, https://doi.org/10.1007/s11081-014-9251-0. (Cited on p. 41.)

[286] T. Schwedes, D. A. Ham, S. W. Funke, and M. D. Piggott, *Mesh Dependence in PDE-Constrained Optimisation*, SpringerBriefs Math. Planet Earth, Springer, Cham, 2017, https://doi.org/10.1007/978-3-319-59483-5. (Cited on p. 54.)

[287] A. Shapiro, *Asymptotic analysis of stochastic programs*, Ann. Oper. Res., 30 (1991), pp. 169–186, https://doi.org/10.1007/BF02204815. (Cited on pp. 69 and 80.)

[288] A. Shapiro, *Asymptotic behavior of optimal solutions in stochastic programming*, Math. Oper. Res., 18 (1993), pp. 829–845, https://doi.org/10.1287/moor.18.4.829. (Cited on p. 80.)

[289] A. Shapiro, *Statistical Inference of Stochastic Optimization Problems*, in Probabilistic Constrained Optimization: Methodology and Applications, S. P. Uryasev, ed., Probabilistic Constrained Optimization, Springer, Boston, MA, 2000, pp. 282–307, https://doi.org/10.1007/978-1-4757-3150-7_16. (Cited on p. 72.)

[290] A. Shapiro, *On Duality Theory of Conic Linear Problems*, in Semi-Infinite Programming: Recent Advances, M. Á. Goberna and M. A. López, eds., Springer, Boston, MA, 2001, pp. 135–165, https://doi.org/10.1007/978-1-4757-3403-4_7. (Cited on p. 2.)

[291] A. Shapiro, *Monte Carlo Sampling Methods*, in Stochastic Programming, Handbooks in Oper. Res. Manag. Sci. 10, Elsevier, 2003, pp. 353–425, https://doi.org/10.1016/S0927-0507(03)10006-0. (Cited on pp. 4, 69, 71, 72, and 101.)

[292] A. Shapiro, *Stochastic programming approach to optimization under uncertainty*, Math. Program., 112 (2008), pp. 183–220, https://doi.org/10.1007/s10107-006-0090-4. (Cited on pp. 1, 71, and 76.)

[293] A. Shapiro, *Distributionally robust stochastic programming*, SIAM J. Optim., 27 (2017), pp. 2258–2275, https://doi.org/10.1137/16M1058297. (Cited on pp. 1, 2, 8, and 9.)

[294] A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on Stochastic Programming: Modeling and Theory*, MOS-SIAM Ser. Optim., SIAM, Philadelphia, PA, 2nd ed., 2014, https://doi.org/10.1137/1.9781611973433. (Cited on pp. 2, 41, 70, 72, 74, 75, 76, 83, 96, 97, 98, 99, 100, 103, and 108.)

[295] A. Shapiro and A. Kleywegt, *Minimax analysis of stochastic problems*, Optim. Methods Softw., 17 (2002), pp. 523–542, https://doi.org/10.1080/1055678021000034008. (Cited on pp. 2, 9, and 41.)

[296] A. Shapiro and A. Nemirovski, *On Complexity of Stochastic Programming Problems*, in Continuous Optimization: Current Trends and Modern Applications, V. Jeyakumar and A. Rubinov, eds., Appl. Optim. 99, Springer, Boston, MA, 2005, pp. 111–146, https://doi.org/10.1007/0-387-26771-9_4. (Cited on pp. 41, 71, and 101.)

[297] R. E. Showalter, *Monotone Operators in Banach Space and Nonlinear Partial Differential Equations*, Math. Surveys Monogr. 49, American Mathematical Society, Providence, RI, 1997. (Cited on pp. 58 and 59.)

[298] A. Sichau and S. Ulbrich, *A Second Order Approximation Technique for Robust Shape Optimization*, in Uncertainty in Mechanical Engineering, H. Hanselka, P. Groche, and R. Platz, eds., Appl. Mech. Mater. 104, Trans Tech Publications, 2012, pp. 13–22, https://doi.org/10.4028/www.scientific.net/AMM.104.13. (Cited on pp. 10 and 41.)

[299] A. Sinha, H. Namkoong, and J. Duchi, *Certifying some distributional robustness with principled adversarial training*. Published as a conference paper at ICLR 2018, 2017, https://arxiv.org/abs/1710.10571. (Cited on pp. 2, 9, and 62.)

[300] A. M.-C. So, *Moment inequalities for sums of random matrices and their applications in optimization*, Math. Program., 130 (2011), pp. 125–151, https://doi.org/10.1007/s10107-009-0330-5. (Cited on pp. 1, 5, 8, 9, 34, 35, 36, and 62.)

[301] D. Sorensen, *Newton's Method with a Model Trust Region Modification*, SIAM J. Numer. Anal., 19 (1982), pp. 409–426, https://doi.org/10.1137/0719026. (Cited on p. 15.)

[302] M. Souto, J. D. Garcia, and Á. Veiga, *Exploiting low-rank structure in semidefinite programming by approximate operator splitting*, 2018, https://arxiv.org/abs/1810.05231. (Cited on p. 38.)

[303] G. Stadler, *Elliptic optimal control problems with L^1-control cost and applications for the placement of control devices*, Comput. Optim. Appl., 44 (2009), pp. 159–181, https://doi.org/10.1007/s10589-007-9150-9. (Cited on pp. 72, 84, 86, 89, and 93.)

[304] S. Steffensen and M. Ulbrich, *A new relaxation scheme for mathematical programs with equilibrium constraints*, SIAM J. Optim., 20 (2010), pp. 2504–2539, https://doi.org/10.1137/090748883. (Cited on p. 10.)

[305] O. Stein, *Bi-Level Strategies in Semi-Infinite Programming*, Nonconvex Optim. Appl. 71, Springer, Boston, MA, 2003, https://doi.org/10.1007/978-1-4419-9164-5. (Cited on p. 10.)

[306] R. J. Stern and H. Wolkowicz, *Indefinite Trust Region Subproblems and Nonsymmetric Eigenvalue Perturbations*, SIAM J. Optim., 5 (1995), pp. 286–313, https://doi.org/10.1137/0805016. (Cited on pp. 9, 16, and 37.)

[307] D. Sun, K.-C. Toh, Y. Yuan, and X.-Y. Zhao, *SDPNAL+: A Matlab software for semidefinite programming with bound constraints (version 1.0)*, Optim. Methods Softw., 35 (2020), pp. 87–115, https://doi.org/10.1080/10556788.2019.1576176. (Cited on p. 38.)

[308] T. Sun, W. Shen, B. Gong, and W. Liu, *A priori error estimate of stochastic Galerkin method for optimal control problem governed by stochastic elliptic PDE with constrained control*, J. Sci. Comput., 67 (2015), pp. 405–431, https://doi.org/10.1007/s10915-015-0091-7. (Cited on pp. 70 and 83.)

[309] M. Talagrand, *Isoperimetry and Integrability of the Sum of Independent Banach-Space Valued Random Variables*, Ann. Probab., 17 (1989), pp. 1546–1570, https://doi.org/10.1214/aop/1176991174. (Cited on pp. 73 and 109.)

[310] P. D. Tao and L. T. H. An, *Lagrangian stability and global optimality in nonconvex quadratic minimization over Euclidean balls and spheres*, J. Convex Anal., 2 (1995), pp. 263–276. (Cited on pp. 9 and 16.)

[311] A. L. Teckentrup, R. Scheichl, M. B. Giles, and E. Ullmann, *Further analysis of multilevel Monte Carlo methods for elliptic PDEs with random coefficients*, Numer. Math., 125 (2013), pp. 569–600, https://doi.org/10.1007/s00211-013-0546-4. (Cited on pp. 4, 108, 120, 121, 123, and 129.)

[312] R. Temam, *Navier-Stokes equations. Theory and Numerical Analysis*, Stud. Math. Appl. 2, North-Holland Publishing Co., Amsterdam, 1977. (Cited on pp. 58 and 60.)

[313] R. Temam, *Navier–Stokes Equations and Nonlinear Functional Analysis*, CBMS-NSF Regional Conf. Ser. in Appl. Math., SIAM, Philadelphia, PA, 2nd ed., 1995, https://doi.org/10.1137/1.9781611970050. (Cited on p. 59.)

[314] H. Tiesler, R. M. Kirby, D. Xiu, and T. Preusser, *Stochastic collocation for optimal control problems with stochastic PDE constraints*, SIAM J. Control Optim., 50 (2012), pp. 2659–2682, https://doi.org/10.1137/110835438. (Cited on pp. 73 and 101.)

[315] F. Tröltzsch, *On Finite Element Error Estimates for Optimal Control Problems with Elliptic PDEs*, in Large-Scale Scientific Computing, I. Lirkov, S. Margenov, and J. Waśniewski, eds., Lecture Notes in Comput. Sci. 5910, Berlin, 2010, Springer, pp. 40–53, https://doi.org/10.1007/978-3-642-12535-5_4. (Cited on p. 70.)

[316] F. Tröltzsch, *Optimal Control of Partial Differential Equations: Theory, Methods and Applications*, Grad. Stud. Math. 112, American Mathematical Society, Providence, RI, 2010. Translated by Jürgen Sprekels. (Cited on pp. 72, 90, and 94.)

[317] F. Tröltzsch and S. Volkwein, *The SQP method for control constrained optimal control of the Burgers equation*, ESAIM Control. Optim. Calc. Var., 6 (2001), pp. 649–674, https://doi.org/10.1051/cocv:2001127. (Cited on p. 3.)

[318] J. A. Tropp, A. Yurtsever, M. Udell, and V. Cevher, *Streaming low-rank matrix approximation with an application to scientific simulation*, SIAM J. Sci. Comput., 41 (2019), pp. A2430–A2463, https://doi.org/10.1137/18M1201068. (Cited on p. 5.)

[319] N.-K. Tsing, M. K. Fan, and E. I. Verriest, *On analyticity of functions involving eigenvalues*, Linear Algebra Appl., 207 (1994), pp. 159–180, https://doi.org/10.1016/0024-3795(94)90009-4. (Cited on pp. 9, 14, 21, and 22.)

[320] B. Turett, *Fenchel–Orlicz spaces*, Dissertationes Mathematicae, Rozprawy Matematyczne 181, Instytut Matematyczny Polskiej Akademii Nauk, Warszawa, 1980, http://pldml.icm.edu.pl/pldml/element/bwmeta1.element.zamlynska-d9dff287-0135-4ff6-91df-a39a66041945. (Cited on p. 110.)

[321] M. Ulbrich, *Semismooth Newton Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces*, MOS-SIAM Ser. Optim., SIAM, Philadelphia, PA, 2011, https://doi.org/10.1137/1.9781611970692. (Cited on pp. 27, 58, 84, and 89.)

[322] M. Ulbrich and B. van Bloemen Waanders, *An introduction to partial differential equations constrained optimization*, Optim. Eng., 19 (2018), pp. 515–520, https://doi.org/10.1007/s11081-018-9398-1. (Cited on p. 42.)

[323] E. Ullmann, H. C. Elman, and O. G. Ernst, *Efficient Iterative Solvers for Stochastic Galerkin Discretizations of Log-Transformed Random Diffusion Problems*, SIAM J. Sci. Comput., 34 (2012), pp. A659–A682, https://doi.org/10.1137/110836675. (Cited on p. 129.)

[324] N. N. Vakhaniya, V. V. Kvaratskheliya, and V. I. Tarieladze, *Weakly Sub-Gaussian Random Elements in Banach Spaces*, Ukrainian Math. J., 57 (2005), pp. 1387–1412, https://doi.org/10.1007/s11253-006-0003-y. (Cited on p. 110.)

[325] A. Van Barel and S. Vandewalle, *Robust Optimization of PDEs with Random Coefficients Using a Multilevel Monte Carlo Method*, SIAM/ASA J. Uncertainty Quantification, 7 (2019), pp. 174–202, https://doi.org/10.1137/17M1155892. (Cited on pp. 41 and 87.)

[326] B. Vexler, *Finite Element Approximation of Elliptic Dirichlet Optimal Control Problems*, Numer. Funct. Anal. Optim., 28 (2007), pp. 957–973, https://doi.org/10.1080/01630560701493305. (Cited on pp. 70 and 80.)

[327] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, *SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python*, Nat. Methods, 17 (2020), pp. 261–272, https://doi.org/10.1038/s41592-019-0686-2. (Cited on pp. 60 and 124.)

[328] S. Volkwein, *Mesh-Independence of an Augmented Lagrangian-SQP Method in Hilbert Spaces and Control Problems for the Burgers Equation*, Dissertation, Technical University of Berlin, Berlin, 1997, https://imsc.uni-graz.at/volkwein/diss.ps. (Cited on pp. 51, 52, 53, 54, 56, 57, 60, and 62.)

[329] S. Volkwein, *Application of the Augmented Lagrangian-SQP Method to Optimal Control Problems for the Stationary Burgers Equation*, Comput. Optim. Appl., 16 (2000), pp. 57–81, https://doi.org/10.1023/A:1008777520259. (Cited on pp. 51 and 54.)

[330] S. Volkwein, *Boundary Control of the Burgers Equation: Optimality Conditions and Reduced-order Approach*, in Optimal Control of Complex Structures, K.-H. Hoffmann, I. Lasiecka, G. Leugering, J. Sprekels, and F. Tröltzsch, eds., Internat. Ser. Numer. Math. 139, Basel, 2001, Birkhäuser, pp. 267–278, https://doi.org/10.1007/978-3-0348-8148-7_22. (Cited on p. 56.)

[331] S. Volkwein, *Second-order conditions for boundary control problems of the Burgers equation*, Control Cybernet., 30 (2001), pp. 249–278, http://yadda.icm.edu.pl/baztech/element/bwmeta1.element.baztech-article-BAT2-0001-1565. (Cited on pp. 57, 58, and 59.)

[332] S. Volkwein, *Lagrange-SQP Techniques for the Control Constrained Optimal Boundary Control for the Burgers Equation*, Comput. Optim. Appl., 26 (2003), pp. 253–284, https://doi.org/10.1023/A:1026047622744. (Cited on p. 56.)

[333] D. Wachsmuth and A. Rösch, *How to Check Numerically the Sufficient Optimality Conditions for Infinite-dimensional Optimization Problems*, in Optimal Control of Coupled Systems of Partial Differential Equations, K. Kunisch, J. Sprekels, G. Leugering, and F. Tröltzsch, eds., Internat. Ser. Numer. Math. 158, Basel, 2009, Birkhäuser, pp. 297–317, https://doi.org/10.1007/978-3-7643-8923-9_17. (Cited on p. 101.)

[334] D. Wachsmuth and G. Wachsmuth, *Necessary conditions for convergence rates of regularizations of optimal control problems*, in System Modeling and Optimization, D. Hömberg and F. Tröltzsch, eds., IFIP Adv. Inf. Commun. Technol. 391, Berlin, 2013, Springer, pp. 145–154, https://doi.org/10.1007/978-3-642-36062-6_15. (Cited on p. 93.)

[335] G. Wachsmuth and D. Wachsmuth, *Convergence and regularization results for optimal control problems with sparsity functional*, ESAIM Control. Optim. Calc. Var., 17 (2011), pp. 858–886, https://doi.org/10.1051/cocv/2010027. (Cited on pp. 72, 84, 89, 90, 91, 92, and 93.)

[336] A. Wächter and L. T. Biegler, *On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming*, Math. Program., 106 (2006), pp. 25–57, https://doi.org/10.1007/s10107-004-0559-y. (Cited on pp. 7, 28, and 29.)

[337] M. J. Wainwright, *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*, Camb. Ser. Stat. Probab. Math., Cambridge University Press, Cambridge, 2019, https://doi.org/10.1017/9781108627771. (Cited on pp. 5 and 100.)

[338] W. Walter, *Ordinary Differential Equations*, Grad. Texts in Math. 182, Springer, New York, NY, 1998, https://doi.org/10.1007/978-1-4612-0601-9. (Cited on p. 59.)

[339] W. Wang and S. Ahmed, *Sample average approximation of expected value constrained stochastic programs*, Oper. Res. Lett., 36 (2008), pp. 515–519, https://doi.org/10.1016/j.orl.2008.05.003. (Cited on pp. 97 and 118.)

[340] Z. Wang, P. W. Glynn, and Y. Ye, *Likelihood robust optimization for data-driven problems*, Comput. Manag. Sci., 13 (2016), pp. 241–261, https://doi.org/10.1007/s10287-015-0240-3. (Cited on pp. 2, 9, and 10.)

[341] M. J. Weinstein and A. V. Rao, *Algorithm 984: ADiGator, a Toolbox for the Algorithmic Differentiation of Mathematical Functions in MATLAB Using Source Transformation via Operator Overloading*, ACM Trans. Math. Softw., 44 (2017), p. 21, https://doi.org/10.1145/3104990. (Cited on p. 29.)

[342] P. Whittle, *Risk-sensitive linear/quadratic/Gaussian control*, Adv. in Appl. Probab., 13 (1981), pp. 764–777, https://doi.org/10.2307/1426972. (Cited on p. 41.)

[343] P. Whittle, *Optimization over Time—Volume 1: Dynamic Programming and Stochastic Control*, John Wiley & Sons, New York, NY, 1982. (Cited on p. 41.)

[344] W. Wiesemann, D. Kuhn, and M. Sim, *Distributionally Robust Convex Optimization*, Oper. Res., 62 (2014), pp. 1358–1376, https://doi.org/10.1287/opre.2014.1314. (Cited on pp. 1, 2, 8, 9, 10, and 40.)

[345] W. A. Woyczyński, *Geometry and Martingales in Banach Spaces*, CRC Press, Boca Raton, FL, 2019. (Cited on pp. 111 and 113.)

[346] H. Xu, Y. Liu, and H. Sun, *Distributionally robust optimization with matrix moment constraints: Lagrange duality and cutting plane methods*, Math. Program., 169 (2018), pp. 489–529, https://doi.org/10.1007/s10107-017-1143-6. (Cited on pp. 2 and 10.)

[347] H. Xu and D. Zhang, *Smooth sample average approximation of stationary points in nonsmooth stochastic optimization and applications*, Math. Program., 119 (2009), pp. 371–401, https://doi.org/10.1007/s10107-008-0214-0. (Cited on p. 69.)

[348] H. K. Xu, *Inequalities in Banach spaces with applications*, Nonlinear Anal., 16 (1991), pp. 1127–1138, https://doi.org/10.1016/0362-546X(91)90200-K. (Cited on p. 112.)

[349] M. Xu, J. J. Ye, and L. Zhang, *Smoothing SQP methods for Solving Degenerate Nonsmooth Constrained Optimization Problems with Applications to Bilevel Programs*, SIAM J. Optim., 25 (2015), pp. 1388–1410, https://doi.org/10.1137/140971580. (Cited on pp. 9, 10, and 27.)

[350] Y. Xu, W. Sun, and L. Qi, *A feasible direction method for the semidefinite program with box constraints*, Appl. Math. Lett., 24 (2011), pp. 1874–1881, https://doi.org/10.1016/j.aml.2011.05.010. (Cited on pp. 9, 12, and 13.)

[351] V. Yakubovich, *Nonconvex optimization problem: The infinite-horizon linear-quadratic control problem with quadratic constraints*, Systems Control Lett., 19 (1992), pp. 13–22, https://doi.org/10.1016/0167-6911(92)90034-P. (Cited on p. 9.)

[352] H. Yamashita and H. Yabe, *A survey of numerical methods for nonlinear semidefinite programming*, J. Oper. Res. Soc. Japan, 58 (2015), pp. 24–60, https://doi.org/10.15807/jorsj.58.24. (Cited on p. 9.)

[353] L. Yang, D. Sun, and K.-C. Toh, *SDPNAL+: a majorized semismooth Newton-CG augmented Lagrangian method for semidefinite programming with nonnegative constraints*, Math. Program. Comput., 7 (2015), pp. 331–366, https://doi.org/10.1007/s12532-015-0082-6. (Cited on p. 38.)

[354] I. Yanikoğlu and D. den Hertog, *Safe approximations of ambiguous chance constraints using historical data*, INFORMS J. Comput., 25 (2013), pp. 666–681, https://doi.org/10.1287/ijoc.1120.0529. (Cited on p. 2.)

[355] V. V. Yurinskiĭ, *On the accuracy of normal approximation of the probability of hitting a ball*, Theory Probab. Appl., 27 (1983), pp. 280–289, https://doi.org/10.1137/1127030. (Cited on p. 109.)

[356] V. Yurinsky, *Sums and Gaussian Vectors*, Lecture Notes in Math. 1617, Springer, Berlin, 1995, https://doi.org/10.1007/BFb0092599. (Cited on pp. 73, 77, 103, 109, 111, 123, 129, and 133.)

[357] M. J. Zahr, K. T. Carlberg, and D. P. Kouri, *An efficient, globally convergent method for optimization under uncertainty using adaptive model reduction and sparse grids*, SIAM/ASA J. Uncertainty Quantification, 7 (2019), pp. 877–912, https://doi.org/10.1137/18M1220996. (Cited on p. 73.)

[358] Y. Zhang, *General Robust-Optimization Formulation for Nonlinear Programming*, J. Optim. Theory Appl., 132 (2007), pp. 111–124, https://doi.org/10.1007/s10957-006-9082-z. (Cited on p. 10.)

[359] C. Zhao and Y. Guan, *Data-driven risk-averse stochastic optimization with Wasserstein metric*, Oper. Res. Lett., 46 (2018), pp. 262–267, https://doi.org/10.1016/j.orl.2018.01.011. (Cited on pp. 2 and 9.)

[360] X.-Y. Zhao, D. Sun, and K.-C. Toh, *A Newton-CG Augmented Lagrangian Method for Semidefinite Programming*, SIAM J. Optim., 20 (2010), pp. 1737–1765, https://doi.org/10.1137/080718206. (Cited on p. 38.)

[361] S. Zymler, D. Kuhn, and B. Rustem, *Distributionally robust joint chance constraints with second-order moment information*, Math. Program., 137 (2013), pp. 167–198, https://doi.org/10.1007/s10107-011-0494-7. (Cited on p. 2.)