

DISTRIBUTIONALLY ROBUST SHAPE AND TOPOLOGY OPTIMIZATION

C. DAPOGNY¹, J. PRANDO² AND B. THIBERT²

¹ *Sorbonne Université, Université Paris Cité, CNRS, Inria, Laboratoire Jacques-Louis Lions, LJLL, F-75005 Paris, France.*

² *Univ. Grenoble Alpes, CNRS, Grenoble INP, LJK, 38000 Grenoble, France.*

ABSTRACT. This article aims to introduce the paradigm of distributional robustness from the field of convex optimization to tackle optimal design problems under uncertainty. We consider realistic situations where the physical model, and thereby the cost function of the design to be minimized depend on uncertain parameters. The probability distribution of the latter is itself known imperfectly, through a nominal law, reconstructed from a few observed samples. The distributionally robust optimal design problem is an intricate bilevel program which consists in minimizing the worst value of a statistical quantity of the cost function (typically, its expectation) when the law of the uncertain parameters belongs to a certain “ambiguity set”. We address three classes of such problems: firstly, this ambiguity set is made of the probability laws whose Wasserstein distance to the nominal law is less than a given threshold; secondly, the ambiguity set is based on the first- and second-order moments of the actual and nominal probability laws. Eventually, a statistical quantity of the cost other than its expectation is made robust with respect to the law of the parameters, namely its conditional value at risk. Using techniques from convex duality, we derive tractable, single-level reformulations of these problems, framed over augmented sets of variables. Our methods are essentially agnostic of the optimal design framework; they are described in a unifying abstract framework, before being applied to multiple situations in density-based topology optimization and in geometric shape optimization. Several numerical examples are discussed in two and three space dimensions to appraise the features of the proposed techniques.

CONTENTS

1. Introduction and Related Works	2
2. Distributionally Robust Optimal Design Problems	4
2.1. Presentation of the abstract distributionally robust optimal design framework	4
2.2. Ambiguity sets based on the Wasserstein distance to a nominal law	5
2.3. Ambiguity sets based on moments	7
2.4. Distributional uncertainties using the conditional value at risk	8
3. Distributionally Robust Density-Based Topology Optimization	11
3.1. Topology optimization of linearly elastic structures	11
3.2. Numerical Strategy	12
3.3. Wasserstein distributionally robust topology optimization under load uncertainty	13
3.4. Moment-based distributionally robust optimization of the compliance of an L-shaped beam	19
3.5. Distributionally robust topology optimization under material uncertainties	22
3.6. Reliability-based optimal design of a bridge using the conditional value at risk	25
4. Distributionally Robust Geometric Shape Optimization	27
4.1. Shape optimization of linear elastic structures	28
4.2. Numerical resolution of shape optimization problems by a level set based mesh evolution strategy	30
4.3. Distributionally robust shape optimization of the stress within a 2d T-shaped beam under load uncertainties	31
4.4. Distributionally robust optimization of the shape of a 3d cantilever beam using a moment-based ambiguity set	32
5. Conclusion and Perspectives	38

Appendix A. Hint of Proof of Proposition 2.1	38
Appendix B. Hint of Proof of Proposition 2.2	39
Appendix C. Hint of Proof of Proposition 2.3	40
References	41

1. Introduction and Related Works

Driven by the need for raw material and energy savings and enabled by the constant progress of high-performance computing, the recent advances in shape and topology optimization now make it possible to optimize devices as diverse as mechanical structures, airfoils and magnetic circuits from the early stages of the design process. However, the systematic deployment of optimal design techniques in industry faces critical challenges, related to their ability to accommodate the requirements of fabrication processes such as milling, casting and additive manufacturing [69, 72, 79, 109, 114] and, closer to the purpose of this article, the inevitable uncertainties over the knowledge of real-life conditions.

Uncertainties in optimal design. Optimal design is about minimizing a cost function depending on the design via its “physical behavior”, under constraints of the same nature. The mathematical models governing this behavior are inherently built from parameters: for instance, the displacement of a mechanical structure depends on the applied loads and on the properties of the constituent material (its Young’s modulus and Poisson’s ratio); likewise, the velocity and pressure within a fluid are influenced by its viscosity and density, the electric and magnetic fields of a wave passing through a photonic device are functions of the operating frequency and the refraction index of the medium, etc. In practice, these physical parameters are often subject to uncertainty, either because they are inaccurately measured (which is typical of the loads applied to a mechanical structure), or because they experience uncontrolled fluctuations over time: changes in the ambient medium, wear, etc. Designing systems that are robust to such uncertainties is crucial, as even minor deviations from the assumed values of the physical parameters during the optimization process can severely degrade the performance of the design. This sharp sensitivity is strikingly illustrated by a situation exhibited in [27], where the best elastic microstructure to withstand a specific loading scenario becomes the worst possible configuration when even infinitesimal perturbations of these loads occur. Understandably enough, the ambition to predict designs with stable performance under uncertain conditions has inspired a thriving literature lately, see for instance [77] and [1] for a recent overview.

Existing approaches for optimal design under uncertainties. Various paradigms can be envisioned to incorporate robustness in an optimal design problem, depending on the assumptions about the uncertain physical parameters of the model.

- Worst-case approaches are usually adopted when no information is available, aside from a maximum bound on their amplitude: one minimizes the worst (i.e. maximum) value of the cost functional over all the possible perturbations. Such problems generally incur an intense computational effort due to their min-max bilevel structure [52]. They are tractable in a few very particular situations; for instance, the quadratic dependence of the compliance of a structure on the applied loads allows to reformulate its minimization under load uncertainty as a semi-definite program [12, 37, 59]. In general, however, one has to resort to heuristic, approximate linearization techniques based on the “smallness” of the maximum amplitude of the uncertainties [3, 53]. In addition to their high computational cost, worst-case approaches are often overly pessimistic, as they result in designs showing poor nominal performances, for the sake of anticipating an unlikely worst-case scenario.
- Probabilistic approaches are naturally preferred in the presence of information about the uncertain parameters, e.g. about their law or some of its moments. Typically, the expected value of the cost functional, its standard deviation, or a weighted sum of both are minimized. Alternatively, so-called reliability-based formulations include constraints on the probability of failure of the system, i.e. on the probability that the cost exceed a certain safety threshold, see for instance [56, 113] about the popular First- and Second-Order Reliability Methods (FORM and SORM) in engineering. Both types of probabilistic frameworks raise the need to calculate integrals over the (high-dimensional)

parameter space. This task can be carried out by expansive Monte Carlo sampling methods [22, 26, 54], accelerated by efficient stochastic collocation techniques and parallelized GPU implementations [75, 76]. A different strategy consists in using the stochastic gradient algorithm to mitigate the number of samples required by the Monte-Carlo strategy, to the expense of an increase in the number of optimization iterations [34, 35, 36, 61, 90], see also [48, 49, 106] for related strategies. Yet another possibility is to construct a reduced-order model for the solution of the boundary-value problem at play: assuming a finite-dimensional structure for the uncertain parameters (which typically stems from a truncated Karhunen-Loève expansion), this solution is approximated by a polynomial chaos expansion [64], or a stochastic Galerkin method is used for its computation [105]. Eventually, in this setting also, linearization techniques based on the “small” amplitude of the uncertainties, are applicable to build deterministic approximations for the considered statistical quantities of the cost function of interest, see e.g. our previous work [4] or [51].

Distributionally robust optimization. Classical probabilistic optimal design formulations themselves suffer from a conceptual shortcoming: in realistic applications, the “true” probability distribution $\mathbb{P}_{\text{true}}$ of the uncertain parameters is unknown; at best, it can be roughly approximated based on (often scarce) observations. Distributionally robust optimization has recently emerged as a sensible paradigm to enforce robustness with respect to uncertainties in a way overcoming this objection. In a nutshell, the worst-case of the expectation (or another statistical quantity, such as the standard deviation) of the cost function is minimized over all the probability laws $\mathbb{Q}$ within a so-called ambiguity set $\mathcal{A}$: $\mathcal{A}$ is made of the laws that are “close” to a given nominal scenario $\mathbb{P}$, which is in practice reconstructed from a few samples. Ambiguity sets of various natures can be thought of. The perhaps most intuitive practice is to define $\mathcal{A}$ as the set of laws $\mathbb{Q}$ lying within a certain distance from $\mathbb{P}$; this distance can be measured in terms of the Kullback-Leibler divergence [60], a maximum mean discrepancy [104], or the Wasserstein distance [42]. Other ambiguity sets are made of laws whose first moments are close to those of $\mathbb{P}$ [39, 83]. We refer to [71, 94, 95] for overviews of distributionally robust optimization, to the book [25] for a focus on machine learning, and to the recent surveys [19, 68].

This recent literature about distributional robustness arises in the rich and structured framework of convex optimization. Unfortunately, optimal design problems often lack such convenient features; moreover, they raise specific difficulties such as the large cost entailed by the numerical evaluations of the objective function and its derivative, as they depend on the solution to one or several physical boundary-value problems. Partly for these reasons, despite its natural character, distributionally robust optimization has been seldom considered in engineering design: beyond our note [33], we are only aware of very few contributions in this direction, dealing with particular situations at that. In [63], a model problem of parametric optimal design is considered, where the uncertain parameter is allowed to take a finite number of values. The experiments conducted in there clearly demonstrate the ability of distributionally robust formulations to predict optimal designs that perform well even in scenarios that are not present in the nominal law. The article [62] considers truss structures in a context where the optimized thickness of the bars is subjected to uncertainty; a semi-definite program is derived to handle the worst value of the failure probability of the system when the moments of the law are “close” from reference values.

Summary of our contributions. This article is the natural sequel of our preliminary note [33]. Inspired by recent ideas in robust convex optimization and optimal transport theory, we develop tractable distributionally robust versions for a wide range of realistic optimal design problems, that are compatible with the particularities of this field, and notably its lack of mathematical structure and the large computational cost of evaluating the optimization criteria at play. More precisely, we consider three situations:

- The ambiguity set $\mathcal{A}$ is the set of probability laws $\mathbb{Q}$ whose Wasserstein distance from the nominal law $\mathbb{P}$ is lower than a certain threshold;
- The ambiguity set $\mathcal{A}$ is made of laws whose first- and second-order moments are close to reference values;
- The statistical quantity which is made robust is no longer the mean value of the cost but its conditional value at risk – a useful notion that allows to handle failure probability constraints.

In the spirit of [14, 110], in all three cases, we rely on suitable entropy-based penalizations, that enable the use of convex duality techniques. Importantly, the proposed methodology is agnostic of the shape and topology optimization framework: it is applicable to both density-based topology optimization and geometric shape

optimization based on the method of Hadamard. To emphasize this genericity, the discussion is framed within a formal, abstract optimal design setting insofar as possible.

Organization of the article. Using an abstract optimal design setting, [Section 2](#) introduces the notion of distributional robustness, as well as the three distributionally robust problems analyzed in the sequel and their tractable reformulations. The next sections are devoted to the application of these concepts in two different optimal design settings: [Section 3](#) considers the popular engineering framework of density-based topology optimization. In there, we notably conduct several numerical experiments to appraise the main features of our formulations. Then, [Section 4](#) deals with geometric shape optimization. A conclusion and a few research perspectives are outlined in [Section 5](#). The article ends with three appendices, devoted to formal sketches of the proofs of the mathematical results underlying our developments.

2. Distributionally Robust Optimal Design Problems

This section presents the three types of distributionally robust optimal design problems considered in this article and details their tractable reformulation allowed by convex duality; it takes place in a formal and abstract setting, that we first introduce.

2.1. Presentation of the abstract distributionally robust optimal design framework

Let $\mathcal{U}_{\text{ad}}$ be a subset of a vector space $\mathcal{H}$, made of admissible designs h . In [Section 3](#), $\mathcal{U}_{\text{ad}} = L^\infty(D, [0, 1])$ is made of the density functions h over a given computational domain $D \subset \mathbb{R}^d$ ($d = 2, 3$). In [Section 4](#), with a slight abuse of this framework, $\mathcal{U}_{\text{ad}}$ is a set of “true” shapes $\Omega \subset \mathbb{R}^d$.

The mathematical model characterizing the physical behavior of a design $h \in \mathcal{U}_{\text{ad}}$ involves parameters; these are gathered in a vector ξ , lying in a compact subset Ξ of a finite-dimensional space $\mathbb{R}^k$. Assuming a perfect knowledge of ξ , the optimal design problem of interest reads:

$$(2.1) \quad \min_{h \in \mathcal{U}_{\text{ad}}} \mathcal{C}(h, \xi),$$

where $\mathcal{C}(h, \xi)$ denotes the cost of a design h when the physical parameters at play equal ξ . For simplicity of the exposition, this formulation omits constraints; as demonstrated by the examples in [Sections 3](#) and [4](#), their treatment does not entail any additional difficulty from the conceptual viewpoint.

In practice, the parameters ξ are only known imperfectly: they are modeled as a k -dimensional random vector $\xi \equiv \xi(\omega)$, where the event variable ω belongs to an abstract probability space $(\mathcal{O}, \mathcal{F}, \mu)$. They are described by their law $\mathbb{P}_{\text{true}}$, that is an element in the set $\mathcal{P}(\Xi)$ of probability measures on Ξ :

For each subset $A \subset \Xi$, $\mathbb{P}_{\text{true}}(A) \in [0, 1]$ = probability that ξ belong to A .

As we have mentioned, a simple and popular method to make the nominal problem [\(2.1\)](#) aware of these uncertainties is to optimize the mean value of the cost functional $\mathcal{C}(h, \xi)$, i.e.:

$$(2.2) \quad \min_{h \in \mathcal{U}_{\text{ad}}} J_{\text{mean}}(h), \text{ where } J_{\text{mean}}(h) = \int_{\Xi} \mathcal{C}(h, \xi) \, d\mathbb{P}_{\text{true}}(\xi).$$

Unfortunately, in realistic applications, the “true” law $\mathbb{P}_{\text{true}}$ of ξ is itself unknown; at best, an estimate $\mathbb{P} \in \mathcal{P}(\Xi)$ is available, which is for instance reconstructed from a few observed samples $\xi^1, \dots, \xi^N \in \Xi$:

$$\mathbb{P} = \frac{1}{N} \sum_{i=1}^N \delta_{\xi^i},$$

where $\delta_{\xi^i} \in \mathcal{P}(\Xi)$ denotes the Dirac mass at $\xi^i \in \Xi$, $i = 1, \dots, N$.

Acknowledging this uncertainty about the law of ξ leads to the distributionally robust version of [\(2.1\)](#), minimizing the worst possible value of the expected cost:

$$(2.3) \quad \min_{h \in \mathcal{U}_{\text{ad}}} J_{\text{dr}}(h), \text{ where } J_{\text{dr}}(h) = \sup_{\mathbb{Q} \in \mathcal{A}} \int_{\Xi} \mathcal{C}(h, \xi) \, d\mathbb{Q}(\xi).$$

This formulation features an ambiguity set $\mathcal{A} \subset \mathcal{P}(\Xi)$, made of laws $\mathbb{Q} \in \mathcal{P}(\Xi)$ which are “close” to $\mathbb{P}$. Depending on the available information about $\mathbb{P}$, the set $\mathcal{A}$ may take different forms. This article proposes three variations around this general concern:

- In [Section 2.2](#), we consider problems where the ambiguity set $\mathcal{A}$ is made of the probability laws $\mathbb{Q}$ which are “close” to $\mathbb{P}$ in terms of (a suitable version of) the Wasserstein distance.
- In [Section 2.3](#), $\mathcal{A}$ is defined as the set of probability laws whose first and second-order moments are “close” to given reference values up to a certain threshold.
- [Section 2.4](#) eventually deals with a slightly different situation, where another measure of risk than the mean value of the cost function is made robust, namely, its conditional value at risk.

Remark 2.1.

- In certain applications, the ambiguity set $\mathcal{A} \equiv \mathcal{A}(h)$ in [\(2.3\)](#) depends on the design h itself, see e.g. [\[40\]](#). In the (scarce) literature devoted to such problems, [\(2.3\)](#) is treated by “freezing” the ambiguity set at the current iteration of the optimization process. Thus, the solution of a version of [\(2.3\)](#) featuring a design-dependent ambiguity set boils down to that of a series of such problems where this set does not depend on h . For this reason, throughout the sequel, we limit ourselves to considering design-independent ambiguity sets $\mathcal{A}$.
- The assumption that the uncertain parameter ξ should lie in a finite-dimensional vector space $\mathbb{R}^k$ may seem restrictive at first glance. In practice, though, infinite-dimensional random vectors are often approximated by a finite-dimensional reduced-order model, obtained e.g. via a truncated Karhunen-Loève expansion, see [Section 3.5](#).

2.2. Ambiguity sets based on the Wasserstein distance to a nominal law

This section considers the situation where the ambiguity set $\mathcal{A}$ in [\(2.3\)](#) is defined as the set of probability laws which are “close” to the nominal law $\mathbb{P}$ in terms of a suitable distance between probability measures, namely, the (entropy-regularized) Wasserstein distance from optimal transport theory. We start with a brief reminder about this notion.

2.2.1. Preliminaries about the Wasserstein distance

Recalling that Ξ is a compact subset of the finite-dimensional space $\mathbb{R}^k$, let $\mathbb{P}, \mathbb{Q}$ be two elements in the set $\mathcal{P}(\Xi)$ of probability measures on Ξ . The Wasserstein distance $W(\mathbb{P}, \mathbb{Q})$ between $\mathbb{P}$ and $\mathbb{Q}$ is defined as the minimum cost of sending the mass of $\mathbb{P}$ onto that of $\mathbb{Q}$, precisely:

$$(2.4) \quad W(\mathbb{P}, \mathbb{Q}) = \min \left\{ \int_{\Xi \times \Xi} c(\xi, \zeta) d\pi(\xi, \zeta), \pi \in \mathcal{P}(\Xi \times \Xi), \pi_1 = \mathbb{P}, \pi_2 = \mathbb{Q} \right\}.$$

This definition involves couplings π , that is, probability measures on the product set $\Xi \times \Xi$; the first and second marginals of such a coupling are respectively denoted by π_1 and π_2 . These are defined through their action on the set $\mathcal{C}(\Xi)$ of continuous functions on Ξ :

$$\forall \varphi \in \mathcal{C}(\Xi), \quad \int_{\Xi} \varphi(\xi) d\pi_1(\xi) = \int_{\Xi \times \Xi} \varphi(\xi) d\pi(\xi, \zeta), \text{ and } \int_{\Xi} \varphi(\zeta) d\pi_2(\zeta) = \int_{\Xi \times \Xi} \varphi(\zeta) d\pi(\xi, \zeta).$$

Intuitively, $\pi(\xi, \zeta)$ represents the amount of mass transferred from ξ to ζ according to the transport plan π , $\pi_1(\xi)$ is the total mass leaving ξ and $\pi_2(\zeta)$ is that arriving at ζ , so that the marginal constraints in [\(2.4\)](#) encode that π transports the “mass distribution” $\mathbb{P}$ onto $\mathbb{Q}$. The “ground cost” $c(\xi, \zeta)$ evaluates the cost of transporting one unit of mass from ξ to ζ ; throughout the article, we use the quadratic cost

$$(2.5) \quad \forall \xi, \zeta \in \Xi, \quad c(\xi, \zeta) = |\xi - \zeta|^2.$$

The Wasserstein distance is a principled notion of distance between probability measures. It has recently aroused much enthusiasm to compare shapes under extraneous formats in imaging [\[46\]](#) and as a loss function between discrete probability distributions in machine learning [\[81\]](#). Notably, it allows to smoothly discriminate probability distributions with different supports, contrary to the Kullback-Leibler divergence. We refer for instance to [\[78, 88, 98, 107\]](#) for mathematical background about optimal transport theory.

Remark 2.2. Strictly speaking, it is the square root of the quantity $W(\mathbb{P}, \mathbb{Q})$ in [\(2.4\)](#) (in the present situation where the ground cost $c(\xi, \zeta)$ is given by [\(2.5\)](#)) that is the distance over $\mathcal{P}(\Xi)$ called Wasserstein distance. For simplicity of the terminology, we abusively refer to $W(\mathbb{P}, \mathbb{Q})$ as the Wasserstein distance.

In our applications, we use a regularized version of the Wasserstein distance. Let us define the entropy $H(\pi)$ of a coupling $\pi \in \mathcal{P}(\Xi \times \Xi)$ by:

$$(2.6) \quad H(\pi) = \begin{cases} \int_{\Xi \times \Xi} \log \frac{d\pi}{d\pi_0} d\pi & \text{if } \pi \text{ is absolutely continuous w.r.t. } \pi_0, \\ +\infty & \text{otherwise.} \end{cases}$$

Following [14], the above definition of $H(\pi)$ features a reference coupling $\pi_0 \in \mathcal{P}(\Xi \times \Xi)$, which is given by:

$$(2.7) \quad \pi_0(\xi, \zeta) = \mathbb{P}(\xi) \otimes \nu_\xi(\zeta), \text{ where } \nu_\xi(\zeta) = \alpha_\xi e^{-\frac{|\xi - \zeta|^2}{2\sigma^2}} \mathbb{1}_\Xi(\zeta),$$

and the normalization constant α_ξ is tailored so that $\nu_\xi \in \mathcal{P}(\Xi)$. More explicitly, the action of π_0 on an arbitrary continuous function $\varphi \in \mathcal{C}(\Xi \times \Xi)$ reads:

$$\int_{\Xi \times \Xi} \varphi(\xi, \zeta) d\pi_0(\xi, \zeta) = \int_{\Xi} \left(\int_{\Xi} \varphi(\xi, \zeta) d\nu_\xi(\zeta) \right) d\mathbb{P}(\xi).$$

In (2.6), a coupling π is called absolutely continuous with respect to π_0 if there exists an integrable function $\alpha \in L^1(\Xi \times \Xi; d\pi_0)$ such that $\pi = \alpha(\xi, \zeta)\pi_0$; the latter is often denoted by $\alpha = \frac{d\pi}{d\pi_0}$.

We eventually define the entropy-regularized Wasserstein distance W_ε via a minimization problem similar to that in (2.4), up to an additional penalization by the entropy of couplings:

$$(2.8) \quad W_\varepsilon(\mathbb{P}, \mathbb{Q}) = \min \left\{ \int_{\Xi \times \Xi} c(\xi, \zeta) d\pi(\xi, \zeta) + \varepsilon H(\pi), \pi \in \mathcal{P}(\Xi \times \Xi), \pi_1 = \mathbb{P}, \pi_2 = \mathbb{Q} \right\}.$$

This addition of a small, strictly convex log-barrier term to the optimal transport problem was proposed in [28], with the purpose to ease numerical computations. Since then, this “blurred” optimal transport has found an interest of its own; for instance, it is often preferred to the original optimal transport formulation in supervised machine learning, where it helps to avoid overfitting. We refer to the lecture notes [86] about entropy-regularized optimal transport.

Remark 2.3. The parameter σ in the definition (2.7) of $\nu_\xi(\zeta)$ reflects the degree of confidence of the user in the nominal law $\mathbb{P}$. Loosely speaking, when σ is “small”, π_0 reduces to the trivial coupling $\mathbb{P}(\xi) \otimes \delta_\xi(\zeta)$, sending $\mathbb{P}$ onto itself by leaving in place each mass unit $\xi \in \Xi$. Then, the entropy penalization $H(\pi)$ in the regularized Wasserstein distance (2.8) imposes that only couplings π very close to this “trivial” one should be considered. When σ gets larger, π_0 acts by “spreading the mass” of $\mathbb{P}$ at each point $\xi \in \Xi$ over a neighborhood with increasing size. The couplings π with “small” entropy values will then have second marginals corresponding to more and more “blurred” versions of $\mathbb{P}$.

Remark 2.4. Contrary to what is suggested by its misleading name, the quantity $W_\varepsilon(\mathbb{P}, \mathbb{Q})$ is solely an approximation of (the square of) a distance. For instance, $W_\varepsilon(\mathbb{P}, \mathbb{P})$ is often different from 0. The function W_ε can be modified into a proper (squared) distance by setting:

$$S_\varepsilon(\mathbb{P}, \mathbb{Q}) := W_\varepsilon(\mathbb{P}, \mathbb{Q}) - \frac{1}{2}W_\varepsilon(\mathbb{P}, \mathbb{P}) - \frac{1}{2}W_\varepsilon(\mathbb{Q}, \mathbb{Q}).$$

This so-called Sinkhorn distance is significantly more complicated to handle; in particular, it does not allow for convenient reformulations of distributionally robust problems similar to those derived below using W_ε . We refer to [46, 100] for more details about the Sinkhorn distance.

2.2.2. Distributionally robust optimization using a Wasserstein ambiguity set

In this section, we study the instance of the distributionally robust optimal design problem (2.3) where the ambiguity set $\mathcal{A}$ is that $\mathcal{A}_W$ defined by:

$$(2.9) \quad \mathcal{A}_W = \left\{ \mathbb{Q} \in \mathcal{P}(\Xi), W_\varepsilon(\mathbb{P}, \mathbb{Q}) \leq m \right\},$$

i.e. $\mathcal{A}_W$ is made of the probability laws $\mathbb{Q} \in \mathcal{P}(\Xi)$ that are within a distance $m > 0$ from the given nominal law $\mathbb{P}$. The distributionally robust problem (2.3) rewrites in this case:

$$(2.10) \quad \min_{h \in \mathcal{U}_{\text{ad}}} J_W(h), \text{ where } J_W(h) = \sup_{\mathbb{Q} \in \mathcal{A}_W} \int_{\Xi} \mathcal{C}(h, \xi) d\mathbb{Q}(\xi).$$

The idea of considering ambiguity sets based on the Wasserstein distance to a fixed nominal law was originally introduced in [89], see [42, 67] for the subsequent developments of this idea and the review [20]. The precise formulation involving the entropy-regularized ambiguity set in (2.8) and (2.9) was proposed in [14, 110] and then reused in our note [33]. We also refer to [13, 47] for an account of its generalization guarantees and to [108] for a recent, open-source implementation of this framework.

The formulation (2.10) is a priori intricate, as it is made of nested minimization and maximization problems. The key point of our methodology is that the worst-case functional $J_W(h)$ can be conveniently reformulated as an infimum, thanks to the following result from convex duality. The latter is proved rigorously in [14]. For completeness, and since its main argument will be reused, a short and formal sketch of the proof is provided in Appendix A.

Proposition 2.1. *Let Ξ be a compact subset of $\mathbb{R}^k$, and let $f : \Xi \rightarrow \mathbb{R}$ be a continuous function. Recalling the definition (2.9) of the ambiguity set $\mathcal{A}_W$, the following formula holds true:*

$$(2.11) \quad \sup_{\mathbb{Q} \in \mathcal{A}_W} \int_{\Xi} f(\xi) d\mathbb{Q}(\xi) = \inf_{\lambda \geq 0} \left(\lambda m + \lambda \varepsilon \int_{\Xi} \log \left(\int_{\Xi} e^{\frac{f(\zeta) - \lambda c(\xi, \zeta)}{\lambda \varepsilon}} d\nu_{\xi}(\zeta) \right) d\mathbb{P}(\xi) \right).$$

Remark 2.5. *As highlighted by the argument in Appendix A, the optimization variable λ in the right-hand minimization problem of (2.11) is the Lagrange multiplier for the inequality constraint $W_{\varepsilon}(\mathbb{P}, \mathbb{Q}) \leq m$ of the maximization problem at the left-hand side. Hence, the optimal point λ^* accounts for the sensitivity of the optimal value of the quantity $\int_{\Xi} f(\xi) d\mathbb{Q}(\xi)$ when the threshold m of this constraint is perturbed. In particular, according to the complementarity relations of the first-order necessary conditions of optimality for (2.11),*

- *If an optimal law $\mathbb{Q}^*$ in the maximization problem at the left-hand side of (2.11) does not saturate the constraint, i.e. if $W_{\varepsilon}(\mathbb{P}, \mathbb{Q}^*) < m$, then $\lambda^* = 0$: the optimal value of $\int_{\Xi} f(\xi) d\mathbb{Q}(\xi)$ does not change if m is slightly modified.*
- *On the contrary, if $W_{\varepsilon}(\mathbb{P}, \mathbb{Q}^*) = m$, then λ^* may differ from 0, and it often does in practice: the optimal value of $\int_{\Xi} f(\xi) d\mathbb{Q}(\xi)$ may change significantly as soon as m is perturbed.*

We refer to e.g. [85] about the classical interpretation of the Lagrange multiplier in optimization theory.

Proposition 2.1 allows to rewrite the distributionally robust problem (2.10) in the following way:

$$(2.12) \quad \min_{(h, \lambda) \in \mathcal{U}_{\text{ad}} \times \mathbb{R}_+} \mathcal{D}_W(h, \lambda) \text{ where } \mathcal{D}_W(h, \lambda) = \lambda m + \lambda \varepsilon \int_{\Xi} \log \left(\int_{\Xi} e^{\frac{c(h, \zeta) - \lambda c(\xi, \zeta)}{\lambda \varepsilon}} d\nu_{\xi}(\zeta) \right) d\mathbb{P}(\xi).$$

This new expression is much more amenable to the numerical treatment than the original one (2.10), as it features a single minimization, over the product space $\mathcal{U}_{\text{ad}} \times \mathbb{R}_+$. Problem (2.12) can be solved owing to standard numerical methods, such as an alternating descent algorithm, see Sections 3 and 4 for illustrations.

Remark 2.6.

- *A result similar to Proposition 2.1 holds when the ambiguity set $\mathcal{A}_W$ is constructed from the classical Wasserstein distance, i.e. when W_{ε} is replaced by W in (2.9). The main difference lies in the replacement of the “log-sum-exp” operation in (2.11) by its formal limit as $\varepsilon \rightarrow 0$, which is a supremum.*
- *Different notions of distance between probability measures could be employed in the definition of the ambiguity set $\mathcal{A}$, such as the Kullback-Leibler divergence, the total variation distance, etc. We refer to [68] for a more exhaustive list of the possibilities and in-depth discussions about the features of these formulations.*

2.3. Ambiguity sets based on moments

This section assumes that the only available information about the law of the uncertain parameter ξ is about its mean value $\mu_0 \in \mathbb{R}^k$ and covariance matrix $\Sigma_0 \in \mathbb{R}^{k \times k}$, as was initially proposed in [39], see also [55, 73]. The ambiguity set $\mathcal{A}_M$ containing the possible probability laws for ξ is defined by:

$$(2.13) \quad \mathcal{A}_M = \left\{ \mathbb{Q} \in \mathcal{P}(\Xi), \left| \int_{\Xi} \xi d\mathbb{Q}(\xi) - \mu_0 \right| \leq m_1, \text{ and } \int_{\Xi} (\xi - \mu_0) \otimes (\xi - \mu_0) d\mathbb{Q}(\xi) \leq m_2 \Sigma_0 \right\}.$$

Here, $m_1 > 0$ and $m_2 > 0$ are given bounds and the second inequality is understood in the sense of positive semi-definite matrices.

In this situation, taking inspiration from the analysis conducted in [Section 2.2](#), we shall derive a tractable alternative form for the distributionally robust problem [\(2.3\)](#) by penalizing the mean value $\int_{\Xi} \mathcal{C}(h, \xi) d\mathbb{Q}(\xi)$ of the cost function by a “small”, strictly convex entropy term. The latter, that we denote by $H(\mathbb{Q})$ (not to be confused with the entropy $H(\pi)$ of couplings in [\(2.6\)](#)), is defined by:

$$(2.14) \quad H(\mathbb{Q}) = \begin{cases} \int_{\Xi} \log \frac{d\mathbb{Q}}{d\mathbb{Q}_0} d\mathbb{Q} & \text{if } \mathbb{Q} \text{ is absolutely continuous w.r.t. } \mathbb{Q}_0, \\ +\infty & \text{otherwise.} \end{cases}$$

Here, the reference law $\mathbb{Q}_0 \in \mathcal{P}(\Xi)$ is (the restriction to Ξ of) the Gaussian distribution with center μ_0 and covariance matrix Σ_0 :

$$\mathbb{Q}_0 = \alpha e^{-\frac{1}{2}(\xi - \mu_0)\Sigma_0^{-1}(\xi - \mu_0)} \mathbf{1}_{\Xi}(\xi) d\xi, \text{ where } \alpha \text{ is a normalization constant.}$$

These ingredients now allow to introduce the following moment-based distributionally robust optimal design problem:

$$(2.15) \quad \min_{h \in \mathcal{U}_{\text{ad}}} J_{\text{M}}(h), \text{ where } J_{\text{M}}(h) := \sup_{\mathbb{Q} \in \mathcal{A}_{\text{M}}} \left(\int_{\Xi} \mathcal{C}(h, \xi) d\mathbb{Q}(\xi) - \varepsilon H(\mathbb{Q}) \right).$$

Again, this a priori intricate problem is made tractable thanks to the following proposition, which converts the supremum functional $J_{\text{M}}(h)$ into an infimum; we refer to [\[93\]](#) for a complete statement and a rigorous proof, a sketch of which is provided in [Appendix B](#).

Proposition 2.2. *Let $\Xi \subset \mathbb{R}^k$ be compact, and let $f : \Xi \rightarrow \mathbb{R}$ be a continuous function. The ambiguity set $\mathcal{A}_{\text{M}}$ being defined by [\(2.13\)](#) with $m_1 > 0, m_2 > 0$, it holds:*

$$(2.16) \quad \sup_{\mathbb{Q} \in \mathcal{A}_{\text{M}}} \left(\int_{\Xi} f(\xi) d\mathbb{Q}(\xi) - \varepsilon H(\mathbb{Q}) \right) = \inf_{\substack{\lambda \geq 0, |\tau| \leq 1, \\ S \in \mathbb{S}_+^k}} \left\{ \lambda m_1 - \lambda \tau \cdot \mu_0 + m_2 S : \Sigma_0 \right. \\ \left. + \varepsilon \log \left(\int_{\Xi} \left(e^{\frac{f(\xi) + \lambda \tau \cdot \xi - S : (\xi - \mu_0) \otimes (\xi - \mu_0)}{\varepsilon}} \right) d\mathbb{Q}_0(\xi) \right) \right\},$$

where $\mathbb{S}_+^k$ is the set of $k \times k$ symmetric and positive semi-definite matrices.

The above proposition paves the way to the following alternative formulation for [\(2.15\)](#):

$$\min_{\substack{h \in \mathcal{U}_{\text{ad}}, |\tau| \leq 1, \\ \lambda \geq 0, S \in \mathbb{S}_+^k}} \mathcal{D}_{\text{M}}(h, \lambda, \tau, S), \text{ where} \\ \mathcal{D}_{\text{M}}(h, \lambda, \tau, S) := \lambda m_1 - \lambda \tau \cdot \mu_0 + m_2 S : \Sigma_0 + \varepsilon \log \left(\int_{\Xi} \left(e^{\frac{\mathcal{C}(h, \xi) + \lambda \tau \cdot \xi - S : (\xi - \mu_0) \otimes (\xi - \mu_0)}{\varepsilon}} \right) d\mathbb{Q}_0(\xi) \right).$$

Again, this minimization problem, posed over an augmented set, can be solved by standard numerical algorithms for optimal design.

2.4. Distributional uncertainties using the conditional value at risk

The previous [Sections 2.2](#) and [2.3](#) have considered distributionally robust functionals accounting for the worst-case value of the expectation of the cost $\mathcal{C}(h, \xi)$ of a design h when the law of ξ belongs to a certain ambiguity set $\mathcal{A}$. Here, we derive robust versions of probabilistic optimal design problems involving a statistical quantity of the cost which is different from its expectation, namely, its conditional value at risk. In spite of its popularity, for instance in mathematical finance, the use of this notion in optimal design seems to be relatively confidential, see however [\[41\]](#) and [\[66\]](#) in the more general context of optimal control of partial differential equations. We exemplify this use in the perspective of reliability-based optimal design, where the conditional value at risk plays the role of a surrogate for a probability of failure.

2.4.1. A few basic facts about value at risk and conditional value at risk

This section collects a few useful facts about the notions of value at risk and conditional value at risk in the context of interest to us; we refer to [96, 97] for details. For the moment, let us assume that the probability law $\mathbb{P} \in \mathcal{P}(\Xi)$ of the parameters ξ of the physical model is known perfectly. For a fixed design $h \in \mathcal{U}_{\text{ad}}$, we denote by $t \mapsto \Psi(h, t)$ the cumulative distribution function of the cost $\mathcal{C}(h, \cdot)$ when the latter is seen as a function of the random parameter $\xi \in \Xi$, that is:

$$\forall t \in \mathbb{R}, \quad \Psi(h, t) = \mathbb{P}\left\{\xi \in \Xi, \mathcal{C}(h, \xi) \leq t\right\}.$$

We recall that the function $t \mapsto \Psi(h, t)$ is non decreasing, with limits:

$$\lim_{t \rightarrow -\infty} \Psi(h, t) = 0, \text{ and } \lim_{t \rightarrow +\infty} \Psi(h, t) = 1.$$

This function is continuous from the right, but it fails to be continuous from the left at those points $t \in \mathbb{R}$ such that $\mathcal{C}(h, \cdot)$ equals t on a subset of Ξ with positive measure, if any [17]. For simplicity, we suppose throughout this presentation that $\Psi(h, \cdot)$ is continuous; although the definitions and intuitions provided below require technical adjustments, the conclusions of our analyses remain valid in the case where it is not.

Let us now define the main objects at stake.

Definition 2.1. *Let $\beta \in [0, 1]$ be a given threshold; then,*

- *The β -value at risk (or β -quantile) $\text{VaR}_\beta(\mathcal{C}(h, \cdot))$ of the cost $\mathcal{C}(h, \cdot)$ is the smallest value $t \in \mathbb{R}$ greater than $\mathcal{C}(h, \cdot)$ with probability β , i.e.*

$$\text{VaR}_\beta(\mathcal{C}(h, \cdot)) = \inf \left\{ t \in \mathbb{R}, \Psi(h, t) \geq \beta \right\}.$$

- *If $\beta < 1$, the β -conditional value at risk $\text{CVaR}_\beta(\mathcal{C}(h, \cdot))$ of $\mathcal{C}(h, \cdot)$ is the average of $\mathcal{C}(h, \cdot)$ over those events where it exceeds the threshold $\text{VaR}_\beta(\mathcal{C}(h, \cdot))$:*

$$(2.17) \quad \text{CVaR}_\beta(\mathcal{C}(h, \cdot)) = \frac{1}{1 - \beta} \int_{\{\xi \in \Xi, \mathcal{C}(h, \xi) \geq \text{VaR}_\beta(\mathcal{C}(h, \cdot))\}} \mathcal{C}(h, \xi) \, d\mathbb{P}(\xi).$$

Note that, by definition of $\text{VaR}_\beta(\mathcal{C}(h, \cdot))$, it holds:

$$(2.18) \quad \mathbb{P}\left\{\xi \in \Xi, \mathcal{C}(h, \xi) \geq \text{VaR}_\beta(\mathcal{C}(h, \cdot))\right\} = 1 - \beta,$$

and so (2.17) is indeed an average. In particular, it holds:

$$(2.19) \quad \text{VaR}_\beta(\mathcal{C}(h, \cdot)) \leq \text{CVaR}_\beta(\mathcal{C}(h, \cdot)).$$

Intuitively, the cost $\mathcal{C}(h, \cdot)$ takes values lower than the β -value at risk $\text{VaR}_\beta(\mathcal{C}(h, \cdot))$ with probability β , see Fig. 1. This quantity is not fully satisfactory, as it does not supply any information about the tail of the distribution of $\mathcal{C}(h, \cdot)$, i.e. those events ξ where $\mathcal{C}(h, \xi)$ exceeds $\text{VaR}_\beta(\mathcal{C}(h, \cdot))$. On the contrary, the conditional value at risk $\text{CVaR}_\beta(\mathcal{C}(h, \cdot))$ is precisely the average of $\mathcal{C}(h, \cdot)$ over these scenarios.

The handling of conditional value at risk is eased by the following representation as a minimum value. The latter is proved in [2, 96]; for the sake of completeness, a formal and intuitive argument is provided in Appendix C.

Proposition 2.3. *The following representation formula holds true:*

$$\text{CVaR}_\beta(\mathcal{C}(h, \cdot)) = \inf_{\alpha \in \mathbb{R}} \left\{ \alpha + \frac{1}{1 - \beta} \int_{\Xi} [\mathcal{C}(h, \xi) - \alpha]_+ \, d\mathbb{P}(\xi) \right\}, \text{ where } [t]_+ := \max(t, 0).$$

The minimum in the above program is uniquely attained at $\alpha = \text{VaR}_\beta(\mathcal{C}(h, \cdot))$.

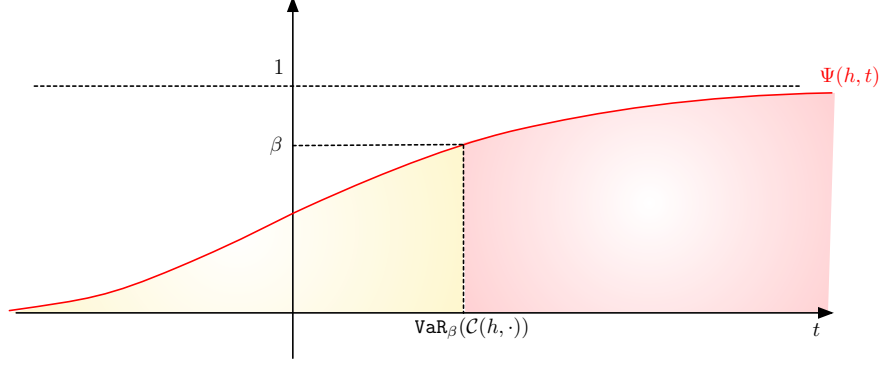


FIGURE 1. The β -value at risk $\text{VaR}_\beta(\mathcal{C}(h, \cdot))$ of the cost $\mathcal{C}(h, \cdot)$ is the lowest level where the cumulative distribution function $\Psi(h, \cdot)$ takes values greater than β ; the β -conditional value at risk $\text{CVaR}_\beta(\mathcal{C}(h, \cdot))$ is the integral of $\mathcal{C}(h, \cdot)$ over those events where it exceeds this value (red area).

2.4.2. Reliability-based optimal design using the conditional value at risk

The notion of conditional value at risk paves the way to a convenient reformulation of optimization problems featuring failure probability constraints, of the form

$$\mathbb{P}\left\{\xi \in \Xi, \mathcal{C}(h, \xi) \geq C_T\right\} \leq 1 - \beta,$$

for a parameter $\beta \in (0, 1)$ and a given safety threshold $C_T \in \mathbb{R}$, see e.g. [24]. Indeed, it follows from the definition of $\Psi(h, t)$ that:

$$\mathbb{P}\left\{\xi \in \Xi, \mathcal{C}(h, \xi) \geq C_T\right\} \leq 1 - \beta \text{ if and only if } \Psi(h, C_T) \geq \beta.$$

Since $\text{VaR}_\beta(\mathcal{C}(h, \cdot))$ is by definition the smallest value $t \in \mathbb{R}$ such that the cost is lower than t with probability β , this is in turn equivalent to:

$$\text{VaR}_\beta(\mathcal{C}(h, \cdot)) \leq C_T.$$

Now invoking (2.19), a conservative surrogate for this constraint is:

$$\text{CVaR}_\beta(\mathcal{C}(h, \cdot)) \leq C_T.$$

Let us apply this observation to a model optimal design problem featuring an objective function $J(h)$ and a constraint on the probability that the cost exceed the threshold C_T :

$$\min_{h \in \mathcal{U}_{\text{ad}}} J(h) \text{ s.t. } \mathbb{P}\left\{\xi \in \Xi \text{ s.t. } \mathcal{C}(h, \xi) \geq C_T\right\} \leq 1 - \beta.$$

The foregoing discussion naturally leads to consider the following problem, featuring a more stringent constraint:

$$\min_{h \in \mathcal{U}_{\text{ad}}} J(h) \text{ s.t. } \text{CVaR}_\beta(\mathcal{C}(h, \cdot)) \leq C_T.$$

Using the representation formula for the Conditional Value at Risk in Proposition 2.3, this problem rewrites as a more conventional optimization problem over the pair (h, α) :

$$(2.20) \quad \min_{\substack{h \in \mathcal{U}_{\text{ad}}, \\ \alpha \in \mathbb{R}}} J(h) \text{ s.t. } \alpha + \frac{1}{1 - \beta} \int_{\xi \in \Xi} [\mathcal{C}(h, \xi) - \alpha]_+ d\mathbb{P}(\xi) \leq C_T.$$

Remark 2.7. In numerical practice, the non smooth function $[\cdot]_+$ featured in (2.20) is replaced by the smooth approximation:

$$[t]_+ \approx \frac{t}{1 + e^{-\gamma t}}, \text{ where } \gamma := 20.$$

2.4.3. Distributionally robust formulation of problems involving the conditional value at risk

Let us now address the situation where the law of the parameter ξ is itself uncertain. To set ideas, we assume that this law belongs to an ambiguity set $\mathcal{A}_W$ of Wasserstein type around a nominal law $\mathbb{P}$, see (2.9). The distributionally robust counterpart of (2.20) then reads:

$$(2.21) \quad \min_{h \in \mathcal{U}_{\text{ad}}} J(h) \text{ s.t. } \sup_{\mathbb{Q} \in \mathcal{A}_W} \min_{\alpha \in \mathbb{R}} \left(\alpha + \frac{1}{1-\beta} \int_{\xi \in \Xi} [\mathcal{C}(h, \xi) - \alpha]_+ d\mathbb{Q}(\xi) \right) \leq C_T.$$

We then use the following proposition, whose proof is similar to that of Proposition 2.1.

Proposition 2.4. *Let Ξ be a compact subset of $\mathbb{R}^k$, and let $f : \Xi \rightarrow \mathbb{R}$ be a continuous function. Recall that $\mathcal{A}_W$ denotes the subset of $\mathcal{P}(\Xi)$ defined by (2.9). The following equality holds true:*

$$\sup_{\mathbb{Q} \in \mathcal{A}_W} \min_{\alpha \in \mathbb{R}} \left(\alpha + \frac{1}{1-\beta} \int_{\xi \in \Xi} f(\xi) d\mathbb{Q}(\xi) \right) = \inf_{\substack{\lambda \geq 0, \\ \alpha \in \mathbb{R}}} \left(\alpha + \frac{\lambda m}{1-\beta} + \frac{\lambda \varepsilon}{1-\beta} \int_{\Xi} \log \left(\int_{\Xi} e^{\frac{f(\zeta) - \lambda c(\xi, \zeta)}{\lambda \varepsilon}} d\nu_{\xi}(\zeta) \right) d\mathbb{P}(\xi) \right).$$

As a result, the distributionally robust optimal design problem (2.21) rewrites:

$$\min_{\substack{h \in \mathcal{U}_{\text{ad}}, \\ \lambda \geq 0, \alpha \in \mathbb{R}}} J(h) \text{ s.t. } \mathcal{D}_C(h, \lambda, \alpha) \leq C_T, \text{ where}$$

$$\mathcal{D}_C(h, \lambda, \alpha) = \alpha + \frac{\lambda m}{1-\beta} + \frac{\lambda \varepsilon}{1-\beta} \int_{\Xi} \log \left(\int_{\Xi} e^{\frac{[\mathcal{C}(h, \zeta) - \alpha]_+ - \lambda c(\xi, \zeta)}{\lambda \varepsilon}} d\nu_{\xi}(\zeta) \right) d\mathbb{P}(\xi).$$

This is a classical constrained optimal design problem, posed over the product space $\mathcal{U}_{\text{ad}} \times \mathbb{R}_+ \times \mathbb{R}$, which can be solved by standard numerical algorithms.

3. Distributionally Robust Density-Based Topology Optimization

This section slips into the framework of density-based optimal design of mechanical structures, about which we refer to the classical book [16], see also [102]. We present various concrete applications of the general methods exposed in Section 2 in this setting. The first Sections 3.1 and 3.2 introduce the physical context and the numerical framework, respectively. Then, Section 3.3 deals with topology optimization problems where the uncertain parameters are the loads applied to the structure, and the ambiguity set $\mathcal{A}$ is based on the Wasserstein distance to a nominal law. The next Section 3.4 illustrates the use of ambiguity sets based on the moments of the uncertain probability law. After that in Section 3.5, we look at another type of uncertainty, concerning the material coefficients. Finally, in Section 3.6, we handle reliability-based optimization problem using the notion of conditional value at risk.

3.1. Topology optimization of linearly elastic structures

Let D be a fixed hold-all domain in $\mathbb{R}^d$ ($d = 2, 3$); the design variable under scrutiny is a density function $h \in L^\infty(D, [0, 1])$, representing a relaxed version of the characteristic function of a subset of D . Specifically,

- The regions where $h(x) = 1$ are made of an isotropic elastic material with Hooke's law A given by:

$$(3.1) \quad \text{For any symmetric } d \times d \text{ matrix } e, \quad Ae = 2\mu e + \lambda \text{tr}(e)I.$$

Here, λ and μ are the Lamé parameters of the material; these are related to the more physical Young's modulus E and Poisson ratio ν via the following relations:

$$(3.2) \quad \mu = \frac{E}{2(1+\nu)}, \text{ and } \lambda = \begin{cases} \frac{E\nu}{(1+\nu)(1-2\nu)} & \text{if } d = 3, \\ \frac{E}{1-\nu^2} & \text{if } d = 2 \text{ (plane stress)}. \end{cases}$$

- The function h equals 0 at points $x \in D$ that are completely surrounded by a very soft material with Hooke's law ηA , $\eta \ll 1$, mimicking void;
- The intermediate regions where $h(x) \in (0, 1)$ are composed of a mixture of material and void: their material properties interpolate between those the bulk and void phases.

These features are incorporated into an inhomogeneous Hooke's tensor $A(h)$ associated to the density function h . According to the Solid Isotropic Material Penalization approach (SIMP), the latter is defined by:

$$(3.3) \quad \forall x \in D, \quad A(h)(x) = \left(\eta + (1 - \eta)h(x)^p \right) A,$$

where the penalization exponent p is typically set to 3.

The considered structures are clamped on a region Γ_D of the boundary ∂D , and external loads $g : \Gamma_N \rightarrow \mathbb{R}^d$ are applied on a disjoint region $\Gamma_N \subset \partial D$. In this situation, the displacement of the structure is the unique solution $u \in H^1(D)^d$ to the system of linear elasticity:

$$(3.4) \quad \begin{cases} -\operatorname{div}(A(h)e(u)) = 0 & \text{in } D, \\ u = 0 & \text{on } \Gamma_D, \\ A(h)e(u)n = g & \text{on } \Gamma_N, \\ A(h)e(u)n = 0 & \text{on } \partial D \setminus \overline{\Gamma_N} \cup \overline{\Gamma_D}. \end{cases}$$

Here, we have denoted by $e(u) := \frac{1}{2}(\nabla u + \nabla u^T)$ the strain tensor associated to a displacement field $u : D \rightarrow \mathbb{R}$. We have also omitted body forces (gravity) for simplicity.

In the sequel, we shall endow the elastic displacement u and the associated physical quantities of interest with various indices or arguments to highlight their dependence on the relevant parameters of this system, notably, the design h , the loads g and the material parameters E, ν .

Remark 3.1. *In practice, the definition of the Hooke's law $A(h)$ in (3.3) involves various filters, imparting effects such as smoothness, among others, which are not mentioned for brevity. Note however that the parameters of these filters are updated in the course of the optimization, which partly explains the large jumps that are sometimes visible in the convergence histories of our numerical results.*

A general topology optimization problem in the above context reads:

$$(3.5) \quad \min_{h \in \mathcal{U}_{\text{ad}}} J(h) \text{ s.t. } G(h) = 0,$$

where the objective and constraint functions $J(h), G(h)$ may depend on h via the elastic displacement u_h . Examples of such functionals are:

- The compliance $C(h)$ of the structure, which equivalently accounts for the elastic energy stored in the structure, or the work done by the external loads:

$$(3.6) \quad C(h) = \int_D A(h)e(u_h) : e(u_h) \, dx = \int_{\Gamma_N} g \cdot u_h \, ds;$$

- The difference between the volume $\operatorname{Vol}(h) = \int_D h \, dx$ of h and a target value V_T .

The differentiation of functions depending on the design via boundary-value problems is a difficult, albeit classical issue in optimal design, see e.g. [16, 23, 92] about the celebrated adjoint method. For completeness, we recall the following derivative formulas, see e.g. [8, 16].

Proposition 3.1. *The compliance and volume functionals are differentiable at an arbitrary density function $h \in \mathcal{U}_{\text{ad}}$, and their derivatives equal:*

$$C'(h)(\hat{h}) = - \int_D A'(h)(\hat{h})e(u_h) : e(u_h) \, dx, \text{ and } \operatorname{Vol}'(h)(\hat{h}) = \int_D \hat{h} \, dx.$$

3.2. Numerical Strategy

Our numerical framework for solving problems of the form (3.5) is based on the open-source implementation [31], and more exactly on a version of the latter dedicated to topology optimization [29]. In a nutshell, it is a steepest-descent strategy based on the derivatives of the objective and constraint functionals $J(h)$ and $G(h)$. The physical boundary-value problems characterizing the elastic displacement u_h in their expressions are solved by the Finite Element Method, using the open-source library **FreeFem** [57]. A descent direction is then inferred from these data thanks to the Null-Space algorithm [44, 45], a constrained optimization algorithm in infinite dimension which is suitable for shape and topology optimization applications, see also [43] for a **python** open-source implementation. These independent operations are orchestrated by a **python** script, and all our computations are performed on a standard **macbook pro** laptop with 16Gb of RAM.

The treatment of distributionally robust problems such as (2.12) raises the need to evaluate the probabilistic integrals over the parameter set Ξ involved in the expressions of the distributionally robust functionals and their derivatives. To achieve this task, we rely on Monte Carlo sampling: every integral of the form $\int_{\Xi} f(\xi) \, d\mathbb{P}(\xi)$ is approximated as:

$$\int_{\Xi} f(\xi) \, d\mathbb{P}(\xi) = \frac{1}{N} \sum_{i=1}^N f(\xi^i),$$

where the ξ^i are N independent and identically distributed samples drawn from the probability distribution $\mathbb{P}$. In practice, in order to keep the computational effort moderate and since our illustrations do not target industrial complexity, we use a maximum number of $N = 10$ vectors; these evaluations can be conducted in parallel thanks to the `numpy` vectorization solutions. On a different note, according to [18], we use the so-called “log-sum-exp trick” to overcome the problems posed by the smallness of the integrand, and thus stabilize computations. Admittedly, this stochastic gradient descent strategy for the resolution of a problem such as (2.12) is overly simplistic, and much more advanced techniques could be used to improve its efficiency, starting from improvements of the sampling strategy by variance reduction techniques, about which we refer to the review article [80], or more advanced versions of the stochastic gradient method, such as that introduced in [48, 49]. Although crucial in real-life applications, we believe that these topics are beyond the scope of the present study.

3.3. Wasserstein distributionally robust topology optimization under load uncertainty

Our first set of numerical examples arises in situations where the loads applied to the structure are uncertain, i.e., $\xi = g$, and the considered cost function is the compliance $C(h, \xi)$ defined in (3.6).

3.3.1. Loads uncertainty with a nominal distribution defined from a single measurement

The structure under scrutiny is a 2d cantilever beam; it is contained in a fixed computational domain D with size 2×1 , which is in practice equipped with a triangular mesh made of about 9,000 vertices ($\approx 17,000$ triangles). The structure is clamped on the left-hand side Γ_D of ∂D , and surface loads are applied on a small region Γ_N near the center of its right-hand side. The corresponding setting is shown in Fig. 2 (a).

In an ideal situation, the structure is subjected to a horizontal load $\xi^0 = (-1, 0)$; we then consider the deterministic optimization problem:

$$(3.7) \quad \min_{h \in \mathcal{U}_{\text{ad}}} C(h, \xi^0) \text{ s.t. } \text{Vol}(h) = V_T,$$

where the target volume V_T equals 0.6. The outcome of its resolution by the algorithm outlined in Section 3.2 is depicted on Fig. 2 (b,c,d). The computation requires 300 iterations, and it lasts about 17 mn.

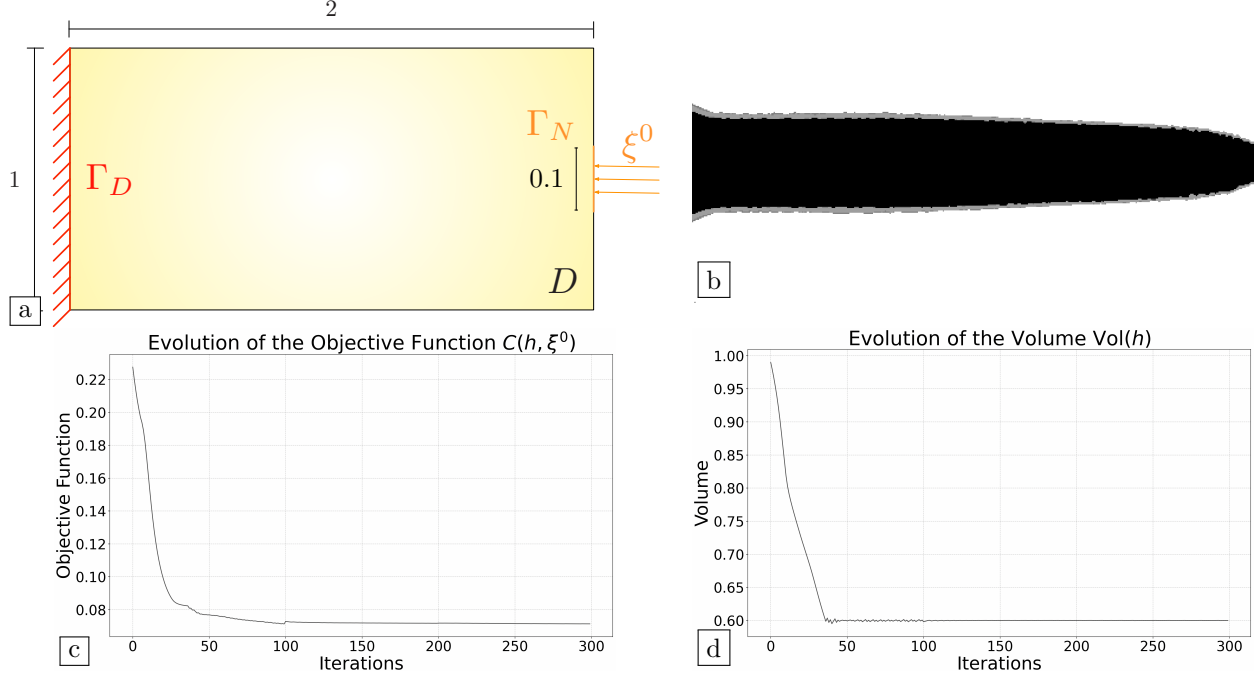


FIGURE 2. (a) Setting of the cantilever beam optimization problem considered in [Section 3.3.1](#); (b) Optimized structure for the ideal problem (3.7); (c) Convergence history of the objective $C(h, \xi^0)$; (d) Convergence history of the volume $\text{Vol}(h)$.

We now assume that the load ξ applied on Γ_N is unknown: it belongs to a sufficiently large ball Ξ in $\mathbb{R}^2$, and its probability law is itself unknown: the latter is estimated solely from the assumed ideal value ξ^0 . Hence, introducing the nominal probability distribution $\mathbb{P} = \delta_{\xi^0}$ made from the single sample ξ^0 , we consider the following Wasserstein distributionally robust problem:

$$(3.8) \quad \min_{h \in \mathcal{U}_{\text{ad}}} \sup_{Q \in \mathcal{A}_W} \int_{\Xi} C(h, \xi) dQ(\xi) \text{ s.t. } \text{Vol}(h) = V_T.$$

Let us recall from the discussion in [Section 2.2.2](#) that this problem has the following tractable equivalent form:

$$(3.9) \quad \min_{\substack{h \in \mathcal{U}_{\text{ad}}, \\ \lambda \geq 0}} \mathcal{D}_W(h, \lambda) \text{ subject to } \text{Vol}(h) = V_T,$$

$$\text{where } \mathcal{D}_W(h, \lambda) := \lambda m + \lambda \varepsilon \log \left(\int_{\Xi} \exp \left(\frac{C(h, \zeta) - \lambda c(\xi^0, \zeta)}{\lambda \varepsilon} \right) d\nu_{\xi}(\zeta) \right),$$

and the probability measure $\nu_{\xi}(\zeta) = \mathbb{1}_{\Xi}(\zeta) \alpha_{\xi} e^{-\frac{|\xi - \zeta|^2}{2\sigma^2}}$ $d\zeta$ is that featured in the definition (2.7) of the reference coupling π_0 . We conduct several numerical investigations in this setting.

Influence of the variance σ^2 and the entropy penalization parameter ε . We solve (3.9) for several values of the parameters σ^2 and ε , for a fixed value $m = 0$ of the maximum bound on the radius of the Wasserstein ball featured in the definition of $\mathcal{A}_W$. Let us point out that the outcome of this analysis is already not trivial since, as noted in [Remark 2.4](#), the quantity $W_{\varepsilon}(\cdot, \cdot)$ is not a true distance, so that the version of (3.9) associated with $m = 0$ may not coincide with the nominal problem (3.7) unless ε tends to 0.

The results of the computation are presented on [Fig. 3](#); each computation involves 600 iterations of our numerical algorithm, for a CPU time of about 2h 45 mn. As expected, when ε and σ^2 are both “very small”, the optimized design is similar to that obtained in the solution of the ideal problem (3.7), displayed on [Fig. 2](#) (b). However, as ε or σ^2 increases, it tends to develop diagonal bars, that are more typical of structures withstanding vertical loads [16]. These observations are in line with the intuitive meanings of

these parameters, discussed in [Section 2.2](#). The variance σ^2 can be interpreted as (the inverse of) the degree of confidence of the user in the nominal law $\mathbb{P}$: for “small” values of σ^2 , the only laws $\mathbb{Q}$ giving rise to a relatively small value of the Wasserstein distance $W_\varepsilon(\mathbb{P}, \mathbb{Q})$ in [\(2.8\)](#) are those very close to $\mathbb{P}$, whereas laws $\mathbb{Q}$ corresponding to increasingly smeared versions of $\mathbb{P}$ are captured in the situation where σ^2 is “not too small”. Likewise, an increase in the value of the entropic regularization parameter ε causes a more and more significant “blurring” of the true Wasserstein distance.

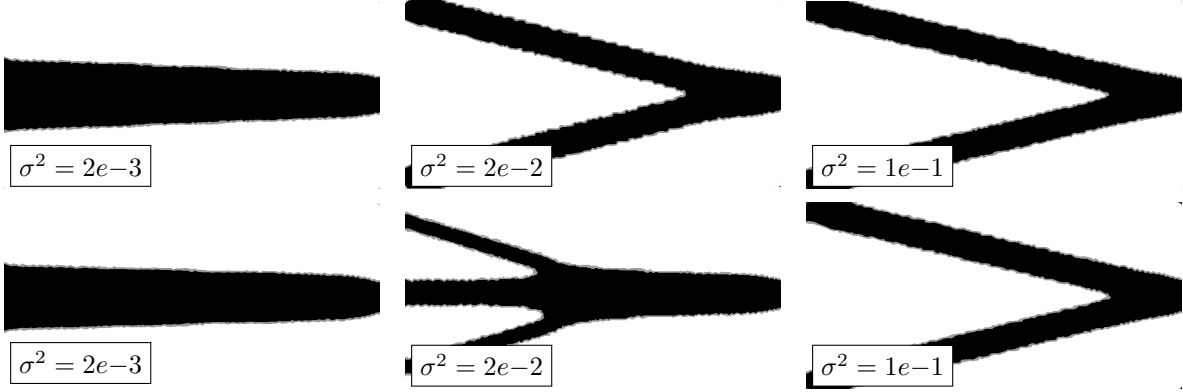


FIGURE 3. Optimized designs of the cantilever beam of [Section 3.3.1](#) obtained by solving the distributionally robust problem [\(3.9\)](#) for $m = 0$ and different values of σ^2 when $m = 0$; (upper row) $\varepsilon = 1e-2$; (lower row) $\varepsilon = 1e-4$.

Influence of the Wasserstein radius m . We now solve [\(3.9\)](#) for several values of the radius m of the Wasserstein ball; the results are depicted on [Fig. 4](#) for $\sigma^2 = 2e-2$, $\varepsilon = 1e-4$, and then $\sigma^2 = 1e-1$, $\varepsilon = 1e-4$.

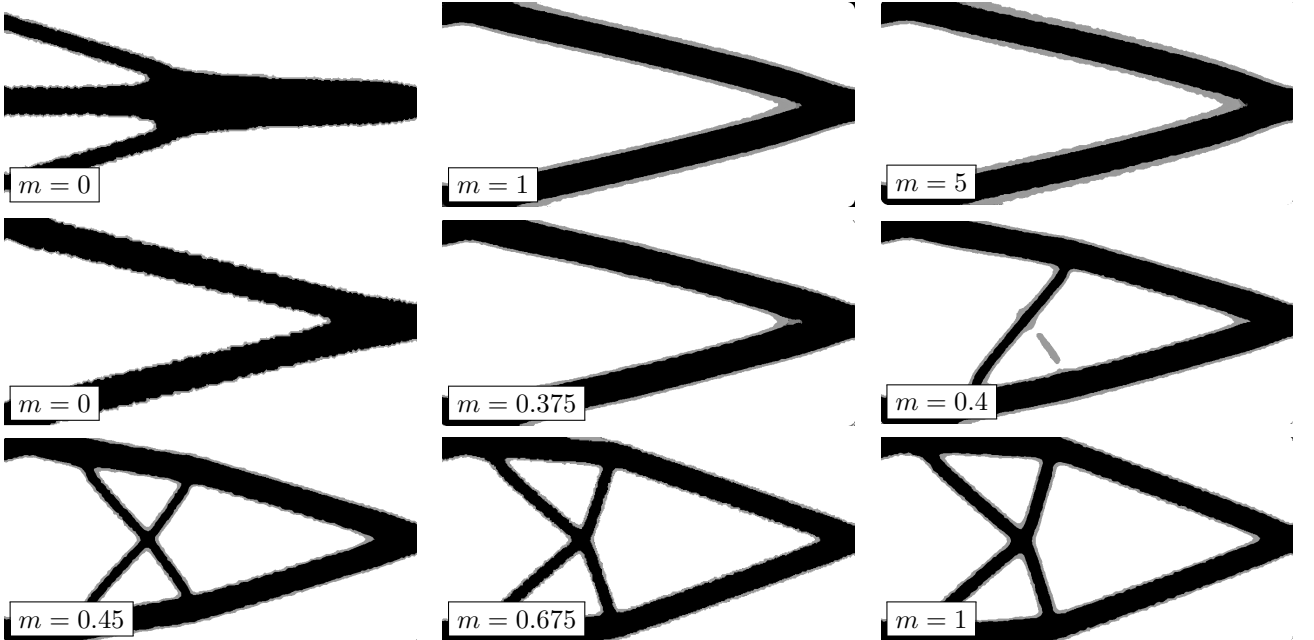


FIGURE 4. Optimized designs of the cantilever beam of [Section 3.3.1](#) obtained by solving the distributionally robust problem [\(3.9\)](#) for several values of the Wasserstein radius m ; (upper row) $\sigma^2 = 2e-2$, $\varepsilon = 1e-4$; (middle and bottom rows) $\sigma^2 = 1e-1$, $\varepsilon = 1e-4$.

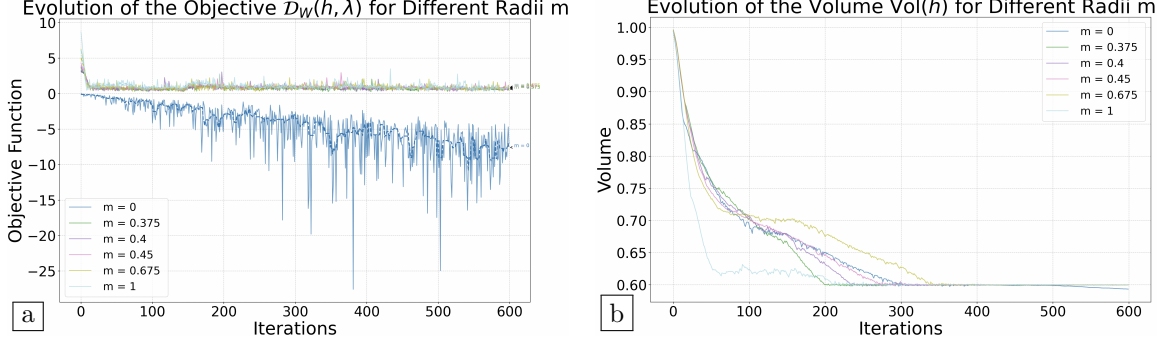


FIGURE 5. *Convergence histories in the distributionally robust optimization of the cantilever considered in Section 3.3.1 for the value $\sigma^2 = 1e-1$ and several values of m ; (a) Objective function; (b) Volume.*

As the Wasserstein radius m increases, the “bad” vertical perturbations of loads are anticipated, urging the formation of diagonal reinforcements that are reminiscent of the typical design of an optimized cantilever beam subjected to vertical loads, see again [16]. The convergence histories of the objective function $\mathcal{D}_W(h, \lambda)$ and of the volume $\text{Vol}(h)$ are presented on Fig. 5 in the case where $\sigma^2 = 1e-1$ and $\varepsilon = 1e-4$. The optimization path is very noisy, but it eventually shows convergence to a local optimum of (3.9), as is typical of a stochastic descent strategy.

Evolution of the Lagrange multiplier λ . Eventually, it is interesting to observe the behavior of the added optimization variable λ in (3.9) with respect to the original formulation (3.8). As noted in Remark 2.5, λ can be interpreted as the Lagrange multiplier for the Wasserstein distance constraint in (3.8), or equivalently as the sensitivity of the maximum value of the mean $\int_{\Xi} \mathcal{C}(h, \xi) d\mathbb{Q}$ with respect to perturbations of the Wasserstein radius m . The evolution of λ is reported on Fig. 6 for various values of the parameters σ^2 , ε and m . Looking first at the case where $\sigma^2 = 2e-2$ and $\varepsilon = 1e-4$ (upper row in Fig. 6), λ keeps increasing through the optimization process when $m = 0$, whereas it rapidly converges to 0 when $m = 1$ or $m = 5$. This is because the Wasserstein distance constraint is saturated in the former case; hence, a perturbation of m is expected to have a large impact on the optimal value in (3.8). On the contrary, this constraint is no longer active in the latter two cases, and so modifying the maximum bound on the Wasserstein distance $W_{\varepsilon}(\mathbb{P}, \mathbb{Q})$ has little to no impact on the optimal value of the objective. Accordingly, we observe that the optimized designs are similar for $m = 1$ and $m = 5$. In the situation where $\sigma^2 = 1e-1$, $\varepsilon = 1e-4$ (lower row in Fig. 6), λ retains positive values when $m = 0$ or $m = 0.25$, which is consistent with the fact seen from Fig. 4 that the optimized shape is still sensitive to the level of the imposed constraint.

3.3.2. Loads uncertainty with a more complex nominal distribution

Our second example illustrates a situation where the nominal law of the uncertain parameter is built from several observations. It concerns the optimal design of a mast, contained in a T-shaped computational domain D with size 2×3 , see Fig. 7 (a). In practice, D is equipped with a mesh composed of approximately 9,000 vertices and 17,000 triangles. The structure is clamped along its bottom boundary Γ_D , and different loads are applied on four disjoint regions collectively denoted by $\Gamma_N = \bigcup_{i=1}^4 \Gamma_{N,i}$. These loads are accordingly described by a vector $\xi = \{\xi_i\}_{i=1,\dots,4}$, contained in a sufficiently large ball $\Xi \subset (\mathbb{R}^2)^4$; ξ is made of four two-dimensional components $\xi_i = (\xi_{i,1}, \xi_{i,2})$, representing the load applied on the part $\Gamma_{N,i}$, for $i = 1, \dots, 4$. More precisely,

- The loads ξ_1 and ξ_2 applied on the arms of the mast represent the weight of the hanging cables.
- The loads ξ_3 and ξ_4 applied on the lower part of the structure model the effect of wind.

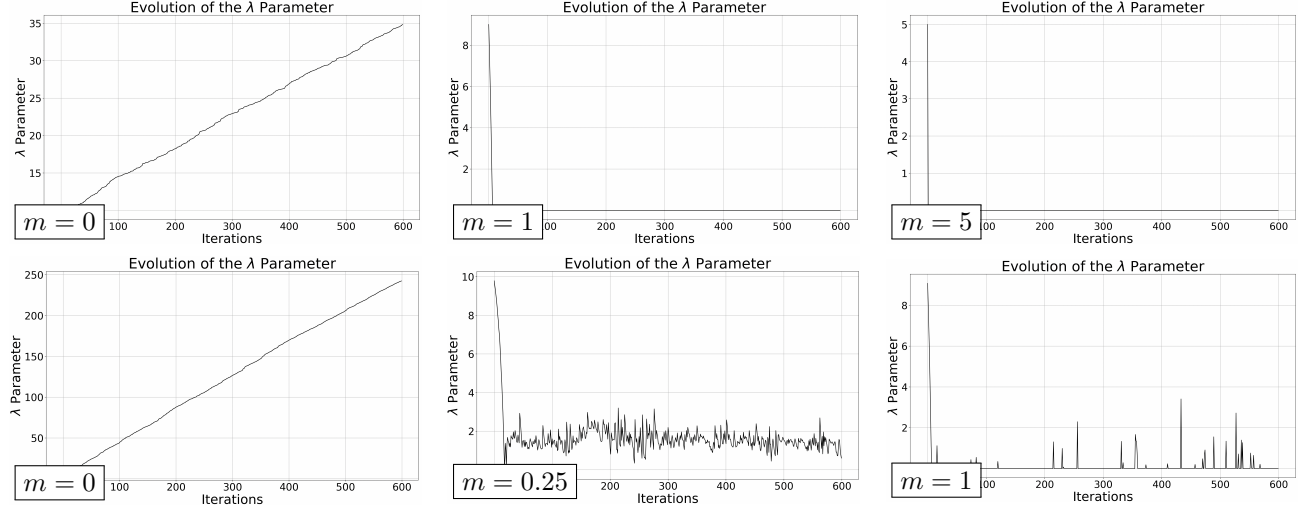


FIGURE 6. Evolution of the Lagrange multiplier λ during the optimization of the cantilever conducted in [Section 3.3.1](#) for various values of the Wasserstein radius m ; (upper row) $\sigma^2 = 2e-2$, $\varepsilon = 1e-4$; (lower row) $\sigma^2 = 1e-1$, $\varepsilon = 1e-4$.

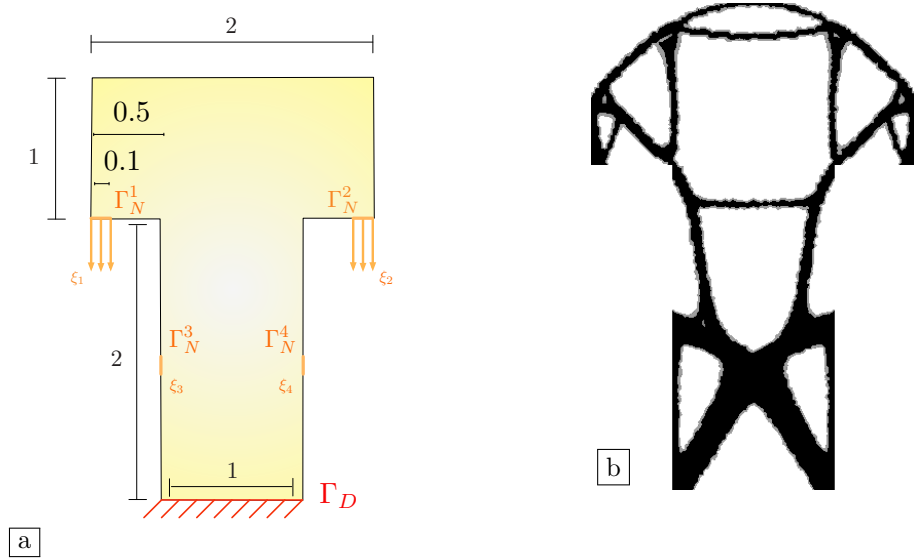


FIGURE 7. (a) Setting of the mast problem considered in [Section 3.3.2](#); (b) Optimized design for the problem (3.12), where the objective function is the mean value of the compliance under the nominal law.

In this context, the displacement of the structure is the solution $u_{h,\xi}$ to the following boundary-value problem:

$$(3.10) \quad \begin{cases} -\operatorname{div}(A(h)e(u)) = 0 & \text{in } D, \\ u = 0 & \text{on } \Gamma_D, \\ A(h)e(u)n = \xi_i & \text{on } \Gamma_{N,i}, \\ A(h)e(u)n = 0 & \text{on } \partial D \setminus \overline{\Gamma_N} \cup \overline{\Gamma_D}. \end{cases}$$

As far as the loads are concerned, we observe three scenarios:

- $\xi^1 = ((0, -1), (0, -1), (0, 0), (0, 0))$, i.e. two vertical loads are applied on the arms $\Gamma_{N,1}$, $\Gamma_{N,2}$ of the mast and the regions $\Gamma_{N,3}$ and $\Gamma_{N,4}$ are free;
- $\xi^2 = ((0, -1), (0, -1), (-1, 0), (-1, 0))$: the same vertical loads are applied on $\Gamma_{N,1}$, $\Gamma_{N,2}$, and $\Gamma_{N,3}$ and $\Gamma_{N,4}$ are subjected to a horizontal force directed to the left, mimicking the effect of wind;
- $\xi^3 = ((0, -1), (0, -1), (1, 0), (1, 0))$: this scenario is similar to the previous one ξ^2 , except that the force applied on $\Gamma_{N,3}$ and $\Gamma_{N,4}$ is directed to the right.

We first empirically attribute the probabilities $\frac{1}{2}$, $\frac{1}{4}$ and $\frac{1}{4}$ to these respective scenarios, so that the law $\mathbb{P} \in \mathcal{P}(\Xi)$ of the uncertain parameter ξ is set to:

$$(3.11) \quad \mathbb{P} = \frac{1}{2}\delta_{\xi^1} + \frac{1}{4}\delta_{\xi^2} + \frac{1}{4}\delta_{\xi^3}.$$

We then optimize the design of the mast with respect to the mean value of its compliance $C(h, \xi)$ defined in (3.6), i.e. we solve:

$$(3.12) \quad \min_{h \in \mathcal{U}_{\text{ad}}} J_{\text{mean}}(h), \text{ s.t. } \text{Vol}(h) = V_T, \text{ where } V_T = 0.3\text{Vol}(D) \text{ and}$$

$$J_{\text{mean}}(h) = \int_{\Xi} C(h, \xi) d\mathbb{P}(\xi) = \frac{1}{2}C(h, \xi^1) + \frac{1}{4}C(h, \xi^2) + \frac{1}{4}C(h, \xi^3).$$

We run 400 iterations of the optimization algorithm outlined in Section 3.2, corresponding to approximately 30 mn of computation; the resulting optimized density function h_{mean}^* is depicted on Fig. 7 (b).

We now come to the situation of interest, where the probability law of the uncertain parameter ξ is uncertain. Having little faith in the empirical reconstruction $\mathbb{P}$ in (3.11), we now minimize the worst value of the expectation of $C(h, \xi)$ when the actual law $\mathbb{Q}$ of ξ is “close” to $\mathbb{P}$ in the sense that it belongs to the Wasserstein ambiguity set $\mathcal{A}_W$ in (2.9):

$$(3.13) \quad \min_{h \in \mathcal{U}_{\text{ad}}} \sup_{\mathbb{Q} \in \mathcal{A}_W} \int_{\Xi} C(h, \xi) d\mathbb{Q}(\xi), \text{ s.t. } \text{Vol}(h) = V_T.$$

Now invoking the material in Section 2.2.2, this problem rewrites:

$$\min_{\substack{h \in \mathcal{U}_{\text{ad}}, \\ \lambda \geq 0}} \mathcal{D}_W(h, \lambda) \text{ s.t. } \text{Vol}(h) = V_T, \quad \text{where}$$

$$\mathcal{D}_W(h, \lambda) = \lambda m + \lambda \varepsilon \int_{\Xi} \log \left(\int_{\Xi} e^{\frac{C(h, \xi) - \lambda c(\xi, \zeta)}{\lambda \varepsilon}} d\nu_{\xi}(\zeta) \right) d\mathbb{P}(\xi),$$

and $\nu_{\xi}(\zeta) \in \mathcal{P}(\Xi)$ is the probability measure (2.7). This example is more involved than the previous one: since the nominal law $\mathbb{P}$ is made of several observations, the objective function $\mathcal{D}_W$ now involves two nested integrals over the parameter space Ξ . In practice, the innermost integral is approximated using 10 samples, so that an evaluation of $\mathcal{D}_W$ requires the solution to 30 versions of (3.10), associated to as many different load cases; in practice, these operations can easily be conducted in parallel. Three designs, $h_{m=0}^*$, $h_{m=5}^*$ and $h_{m=10}^*$, optimized for the respective values $m = 0, 5$ and 10 of the Wasserstein radius, are displayed in Fig. 8. Each computation takes about 600 iterations of our numerical strategy, for a computational time of about 10h.

The three optimized designs resulting from this experiment present less obvious visual differences than in the example of Section 3.3.1. This is because, in the present situation, contrary to the previous one, the three loads ξ^1 , ξ^2 , ξ^3 featured by the nominal law are already quite harmful for the structure compliance, and so the nominal distribution is already not that far from capturing the worst-case situation. However, we observe that as m increases, the arches near the arms of the mast h_{mean}^* turn to flat connections, in response to the anticipation of possible horizontal perturbations of the loads in there. Moreover, the regions of the structure near $\Gamma_{N,3}$ and $\Gamma_{N,4}$ become thicker, in order to better withstand vertical perturbations of the loads.

Out-of-sample analysis. In order to compare the design h_{mean}^* resulting from the minimization (3.12) of the expected compliance with the distributionally robust designs $h_{m=0}^*$, $h_{m=5}^*$ and $h_{m=10}^*$, we compute their compliance $C(h, \xi)$ for several load scenarios; the results are registered in Table 1. The first three scenarios ξ^1 , ξ^2 , ξ^3 are those present in the nominal distribution; the four next vectors $(\xi^i)_{4 \leq i \leq 7}$ model situations where,

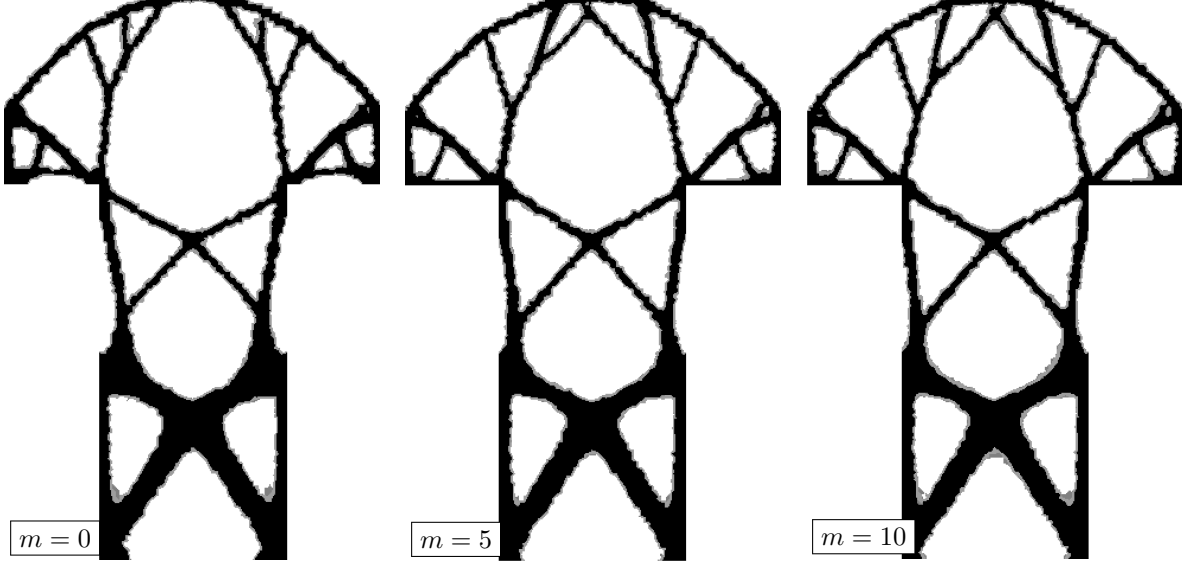


FIGURE 8. Distributionally robust optimized designs $h_{m=0}^*$, $h_{m=5}^*$ and $h_{m=10}^*$ of the electric mast considered in [Section 3.3.2](#) associated to the values $m = 0, 5$ and 10 .

in addition to the weight of the cables hanging from the arms, horizontal stretching or compression effects are added. The final ten vectors $(\xi^i)_{8 \leq i \leq 17}$ have random entries, drawn in the set $\{-1, 1, 0\}$. The design h_{mean}^* optimized for the expectation of the compliance performs best in the first 3 load situations, which is expected since they are present in the nominal law $\mathbb{P}$; on that contrary, h_{mean}^* shows poor performance in the other situations. In contrast, the distributionally robust designs $h_{m=0}^*$, $h_{m=5}^*$ and $h_{m=10}^*$ show lesser performance in the first 3 events, but their efficiency remains stable in the case of loads that are not represented in the nominal distribution. This confirms the ability of the distributionally robust approach to anticipate events that are not present in the nominal distribution, but “close” to some of them, see again [\[63\]](#) for an account of this capability. Note that, in some scenarios, the designs optimized at larger Wasserstein radius have better performance; this happens when the applied load corresponds to a perturbation which is “far” from those in the nominal law, and is thus only effectively captured when the radius of the Wasserstein constraint is large enough. On the contrary, when such perturbations are already captured at the initial radius, increasing the radius further may decrease the impact of this capture, ultimately resulting in lower performance.

3.4. Moment-based distributionally robust optimization of the compliance of an L-shaped beam

This section discusses a situation where the ambiguity set $\mathcal{A} \equiv \mathcal{A}_M$ used in the formulation of the distributionally robust problem is defined from the moments of probability distributions, see [\(2.13\)](#) in [Section 2.3](#).

As illustrated on [Fig. 9 \(a\)](#), we now consider beams, contained in an L-shaped computational box D with dimensions 1×1 ; the domain D is equipped with a mesh comprising approximately 8,500 vertices and 17,000 triangles. The optimized structures are clamped on the upper side Γ_D of ∂D , and a load ξ is applied on a small region Γ_N at its right-hand side.

In ideal circumstances, ξ is known perfectly: it is the unit horizontal vector $\xi^0 = (-1, 0)$. We then solve the problem:

$$(3.14) \quad \min_{h \in \mathcal{U}_{\text{ad}}} C(h, \xi^0) \text{ s.t. } \text{Vol}(h) = V_T.$$

Here, $C(h, \xi)$ stands for the compliance [\(3.6\)](#) of the beam when it is subjected to the load ξ , and the target volume equals $V_T = 0.2$. About 600 iterations of the optimization algorithm sketched in [Section 3.2](#) are applied to this end; the resulting density h_{det}^* is depicted on [Fig. 9 \(b\)](#).

$\xi \backslash h$	h_{mean}^*	$h_{m=0}^*$	$h_{m=5}^*$	$h_{m=10}^*$
$\xi^1 = ((0, -1), (0, -1), (0, 0), (0, 0))$	1.45263	1.58169	1.80382	1.84381
$\xi^2 = ((0, -1), (0, -1), (1, 0), (1, 0))$	3.96816	4.54952	4.95328	5.04692
$\xi^3 = ((0, -1), (0, -1), (-1, 0), (-1, 0))$	3.97635	4.62913	4.93366	5.02842
$\xi^4 = ((0, -1), (0, -1), (-1, 0), (1, 0))$	2.30295	2.56223	2.99777	3.02278
$\xi^5 = ((0, -1), (0, -1), (1, 0), (-1, 0))$	2.0726	1.99442	2.18521	2.21876
$\xi^6 = ((-1, -1), (-1, -1), (-1, 0), (-1, 0))$	90.9951	18.2647	16.5193	16.5381
$\xi^7 = ((1, -1), (1, -1), (-1, 0), (-1, 0))$	81.7526	8.31552	6.92599	6.76357
$\xi^8 = ((1, -1), (1, 1), (-1, -1), (1, 0))$	53.2508	15.6051	10.7477	10.1949
$\xi^9 = ((1, 1), (1, 1), (-1, 1), (0, -1))$	87.6837	12.8845	11.0046	10.9336
$\xi^{10} = ((0, 0), (0, 0), (-1, -1), (0, 0))$	2.55941	2.30058	2.22696	2.22671
$\xi^{11} = ((-1, -1), (-1, 1), (1, 1), (0, 1))$	196.139	28.1766	19.6703	19.3135
$\xi^{12} = ((0, 0), (1, 0), (1, 1), (-1, 1))$	29.1637	6.60941	5.6305	5.56953
$\xi^{13} = ((-1, 1), (1, -1), (-1, 1), (0, -1))$	59.6293	18.1618	12.9192	12.609
$\xi^{14} = ((0, 0), (0, -1), (0, -1), (0, 1))$	11.3029	3.96029	2.79272	2.72636
$\xi^{15} = ((1, -1), (1, 1), (0, -1), (0, 0))$	51.8462	14.3922	9.50305	8.9527
$\xi^{16} = ((0, 0), (0, 0), (0, -1), (1, 0))$	1.51937	1.35681	1.40098	1.40919
$\xi^{17} = ((0, 0), (0, 0), (0, 1), (1, 0))$	1.89735	1.81272	1.77383	1.77781

TABLE 1. Values of the compliance $C(h, \xi)$ of the optimized masts in [Section 3.3.2](#), associated to several values of the load parameter ξ .



FIGURE 9. Optimal design example of an L-shaped beam considered in [Section 3.4](#); (a) Setting of the test-case; (b) Optimized design h_{det}^* in the unperturbed situation; (c) Optimized design h_{mean}^* for the mean value of the compliance $C(h, \xi)$ under the nominal law $\mathbb{P}$.

We now assume that the parameter ξ is uncertain, but that its probability law $\mathbb{P}$ is perfectly known: ξ is a Gaussian vector whose first- and second-order moments μ_0 and Σ_0 are given by:

$$\mu_0 = \xi^0, \text{ and } \Sigma_0 = \sigma^2 \mathbf{I}, \quad \sigma^2 = 1e-2.$$

We then minimize the mean value of the cost functional with respect to $\mathbb{P}$, i.e. we solve:

$$(3.15) \quad \min_{h \in \mathcal{U}_{\text{ad}}} \int_{\Xi} C(h, \xi) \, d\mathbb{P}(\xi), \quad \text{s.t.} \quad \text{Vol}(h) = V_T.$$

Again applying 600 iterations of our algorithm ($\approx 1\text{h}$ of computation), the optimized density h_{mean}^* is that depicted in Fig. 9 (c).

We now turn to a situation where the probability law of ξ is uncertain, and only its first- and second-order moments are known. Following Section 2.3, we then consider the following distributionally robust counterpart of (3.14) and (3.15):

$$\min_{h \in \mathcal{U}_{\text{ad}}} \left(\sup_{\mathbb{Q} \in \mathcal{A}_M} \int_{\Xi} C(h, \xi) \, d\mathbb{Q}(\xi) - \varepsilon H(\mathbb{Q}) \right) \quad \text{s.t.} \quad \text{Vol}(h) = V_T,$$

featuring the moment-based ambiguity set $\mathcal{A}_M$ defined in (2.13); the entropy $H(\mathbb{Q})$ of a law $\mathbb{Q} \in \mathcal{P}(\Xi)$ is defined by (2.14) and the reference law $\mathbb{Q}_0$ in this formula is $\mathbb{Q}_0 = \mathbb{P}$. The parameter ε is set to $\varepsilon = 0.01$. According to Section 2.3, this problem equivalently rewrites:

$$\min_{\substack{h \in \mathcal{U}_{\text{ad}}, |\tau| \leq 1, \\ \lambda \geq 0, S \in \mathbb{S}_+^m}} \mathcal{D}_M(h, \lambda, \tau, S), \quad \text{s.t.} \quad \text{Vol}(h) = V_T \quad \text{where} \\ \mathcal{D}_M(h, \lambda, \tau, S) := \lambda m_1 - \lambda \tau \cdot \mu_0 + m_2 S : \Sigma_0 + \varepsilon \log \left(\int_{\Xi} \left(e^{\frac{C(h, \xi) + \lambda \tau \cdot \xi - S : (\xi - \mu_0)}{\varepsilon}} \right) \, d\mathbb{P}(\xi) \right).$$

We solve this problem for several values of the bounds m_1, m_2 defining the ambiguity set $\mathcal{A}_M$; the optimized designs h_{m_1, m_2}^* are presented in Fig. 10. Each computation requires 600 iterations of our algorithmic strategy, for a CPU time of about 70 mn.

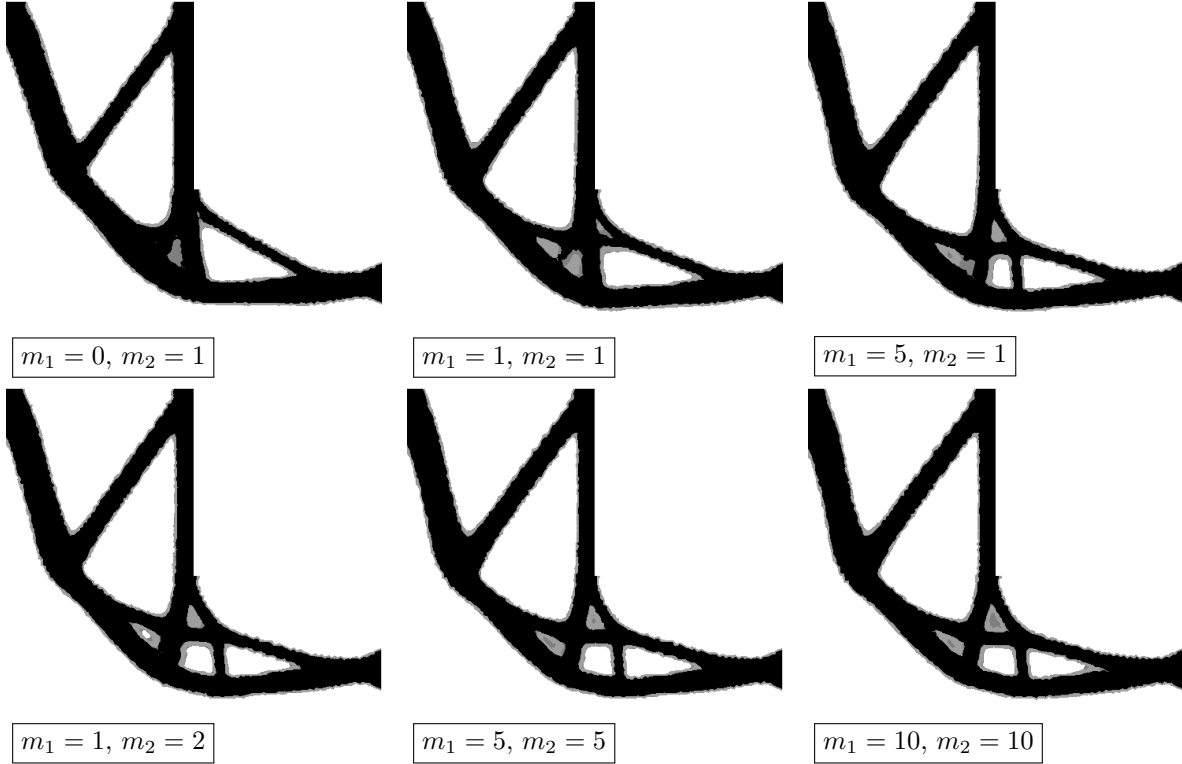


FIGURE 10. Distributionally robust optimized designs obtained in the L-shaped beam example of Section 3.4, associated to various values of the parameters m_1, m_2 of the moment-based ambiguity set $\mathcal{A}_M$.

According to these results, as the ambiguity set gets wider, the optimized designs tend to develop a rounded re-entrant corner. This feature is well-known to better attenuate the stress concentrations induced by vertical loads in this region which is constrained by the geometry of the design domain D , see e.g. [9]. As expected, the compliance $C(h_{m_1, m_2}^*, \xi^0)$ of these designs in the situation where the ideal load ξ^0 is applied deteriorates as the ambiguity set increases, as shown in Table 2. As far as the effect of loads different from ξ^0 is concerned, an out-of-sample analysis in the spirit of that conducted in the previous Section 3.3.2, yields similar conclusions. Omitting full details for brevity – we refer to the thesis [93] about these –, let us solely mention that again, the designs h_{det}^* and h_{mean}^* are less efficient than those resulting from the distributionally robust formulation when they are submitted to a load which is different from ξ^0 . This happens even for loads that are “close” to ξ^0 , although the difference gets of course much more significant when they get farther. This observation again underlines the ability of distributionally robust formulations to accommodate uncertain events that are not part of the observations used for reconstructing the nominal law $\mathbb{P}$.

	h_{det}^*	h_{mean}^*	$h_{m_1=0, m_2=1}^*$	$h_{m_1=1, m_2=1}^*$	$h_{m_1=5, m_2=1}^*$	$h_{m_1=1, m_2=2}^*$	$h_{m_1=5, m_2=5}^*$	$h_{m_1=10, m_2=10}^*$
$C(h, \xi^0)$	0.5409451	0.595818	0.63298	0.661997	0.666666	0.645945	0.659693	0.670644

TABLE 2. Values of the compliance $C(h, \xi^0)$ of the distributionally robust designs of the L-shaped beam considered in Section 3.4 when the nominal load ξ^0 is applied.

3.5. Distributionally robust topology optimization under material uncertainties

The present section considers uncertainties about the properties of the constituent material of the design. This situation departs from Sections 3.3 and 3.4 on at least two aspects: on the one hand, the uncertain parameter ξ does not naturally pertain to a finite-dimensional space; rather, it stems from a truncated Karhunen-Loève expansion of an infinite-dimensional random field. On the other hand, the considered optimal design problem features a cost function which is not self-adjoint, see Proposition 3.2 below.

The designs under scrutiny are gripping mechanisms, see Fig. 12 (a); they are enclosed in a fixed computational domain D with size 1×1 , equipped with a mesh composed of 9,000 vertices ($\approx 17,000$ triangles). A load $g = (0.1, 0)$ is applied on a region Γ_N of the left-hand side of ∂D .

The uncertainties over the physical model concern the (inhomogeneous) Young’s modulus E of the constituent elastic material. Following [50, 70], this function is assumed to be of the form:

$$(3.16) \quad E(x, \xi) = G(\tilde{E}(x, \xi)), \text{ where } \tilde{E}(x, \xi) = E_0(x) + \sum_{i=1}^k \sqrt{\lambda_i} E_i(x) \xi_i, \quad x \in D, \quad \xi \in \Xi.$$

Here,

- The $E_i(x)$ are given, deterministic elements in $L^2(D)$;
- The uncertain parameter ξ lies in a compact subset Ξ of $\mathbb{R}^k$. In the nominal situation, its law $\mathbb{P}$ is the (restriction to Ξ of the) normalized k -variate Gaussian law.
- The function G equals $G := F^{-1} \circ \Phi$, where Φ is the cumulative distribution function of the k -variate Gaussian law and F is the cumulative distribution function of each real-valued random variable $E(x, \cdot)$, $x \in D$. In practice, F is the cumulative distribution function of the uniform law on $[0.1, 1.9] \subset \mathbb{R}$; in particular, $E(x, \xi)$ is uniformly bounded away from 0 and ∞ .

Let us briefly explain how the structure (3.16) of $\tilde{E}$ shows up in practice, see e.g. [26, 51, 70]. The quantity $\tilde{E}$ is modeled as a random field $\tilde{E}(x, \omega)$, where the event variable ω belongs to the abstract probability space $(\mathcal{O}, \mathcal{F}, \mu)$: intuitively, a random variable $\tilde{E}(x, \cdot)$ is attached at each spatial position $x \in D$. The available information about $\tilde{E}(x, \omega)$ concerns its mean value $E_0(x)$ and covariance kernel $\text{Cov}(x, y)$, defined by:

$$\forall x, y \in D, \quad E_0(x) = \int_{\mathcal{O}} \tilde{E}(x, \omega) \, d\mu(\omega), \text{ and } \text{Cov}(x, y) = \int_{\mathcal{O}} (\tilde{E}(x, \omega) - E_0(x))(\tilde{E}(y, \omega) - E_0(y)) \, d\mu(\omega).$$

Here, we assume that $E_0(x) \equiv 1$ and that $\text{Cov}(x, y)$ is given by:

$$(3.17) \quad \forall x, y \in D, \quad \text{Cov}(x, y) = \alpha e^{-\frac{|x-y|}{l_{\text{cor}}}}, \quad \text{where } l_{\text{cor}} = 2e-2 \text{ and } \alpha = 100.$$

The finite-dimensional structure of $\tilde{E}(x, \xi)$ in (3.16) now follows from a Karhunen-Loève decomposition of the covariance kernel $\text{Cov}(x, y)$, truncated after the k^{th} term, see e.g. [74, 99]: the pairs (λ_i, E_i) are the eigenlements of the operator

$$T : L^2(D) \rightarrow L^2(D), \quad T\varphi(x) = \int_D \text{Cov}(x, y)\varphi(y) \, dy.$$

The first six eigenfunctions of (3.17) are displayed on Fig. 11, together with a few realizations of $E(x, \xi)$.

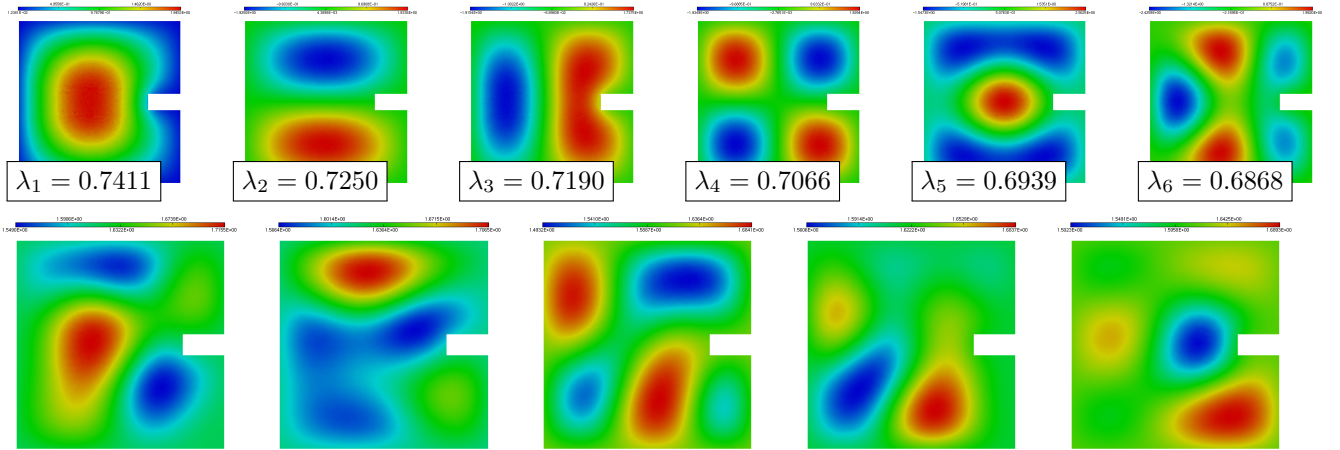


FIGURE 11. *Distributionally robust optimization under material uncertainties: (Upper row) First six eigenpairs (E_i, λ_i) of the covariance kernel (3.17); (Lower row) Realizations of the random field $E(x, \xi)$ in (3.16), associated to five different instances of the parameter ξ .*

Remark 3.2. *In practice, the covariance kernel $\text{Cov}(x, y)$ of the random field $\tilde{E}(x, \omega)$ is itself unknown and it has to be reconstructed from (a few) observed instances of $E(x, \cdot)$ (or $\tilde{E}(x, \cdot)$). Hence, in principle, one should take into account the fact that the basis functions $E_i(x)$ in the decomposition (3.16) are themselves uncertain. This topic of distributionally robust covariance estimators, which goes beyond the scope of this article, is discussed in [38, 84, 112].*

We denote by $u_{h, \xi}$ the solution to the linear elasticity system (3.4) when the Young's modulus featured in the Hooke's law $A \equiv A(h, \xi)$ in (3.1) is that $E(x, \xi)$ given in (3.16), see (3.2). In this situation, we aim to optimize the design h of the structure so that it adopts a prescribed displacement u_T on a region Γ_T located near the jaws of the mechanism. To achieve this, we consider the following problem:

$$(3.18) \quad \min_h D(h, \xi), \quad \text{where } D(h, \xi) = \int_D \chi(x) |u_{h, \xi} - u_T|^2 \, dx$$

and $\chi(x)$ is a given weight factor, which equals 1 in a thin region around the jaws (depicted in blue in Fig. 12 (a)) and 0 elsewhere.

The derivative of the least-square functional $D(h, \xi)$ is given by the following proposition, whose classical proof is omitted for brevity, see [16].

Proposition 3.2. *The function $D(h, \xi)$ is differentiable with respect to the design variable h , and its derivative reads:*

$$D'(h, \xi)(\hat{h}) = \int_D A'(h)(\hat{h}) e(u_{h, \xi}) : e(p_{h, \xi}) \, dx,$$

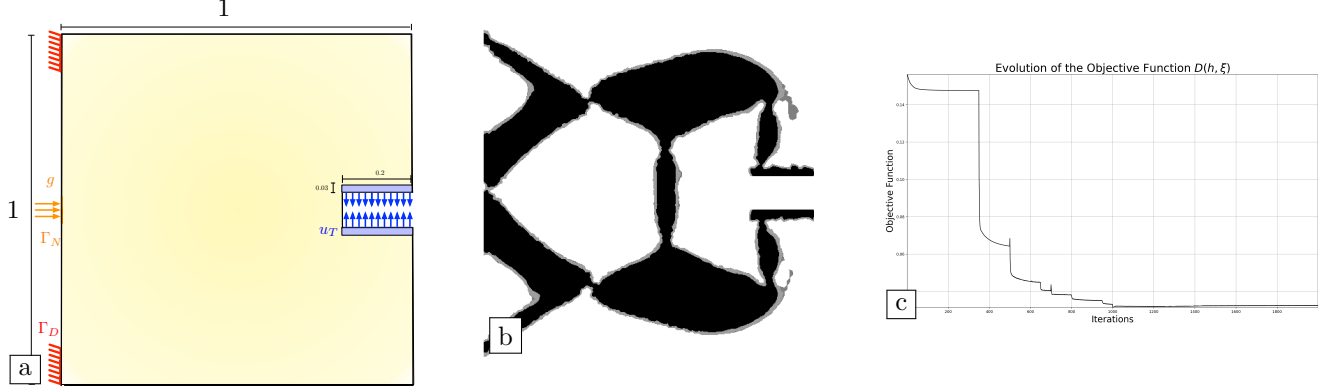


FIGURE 12. (a) Setting of the optimal design example of a gripping mechanism considered in Section 3.5; (b) Optimized design in the ideal situation; (c) Convergence history.

where the adjoint state $p_{h,\xi}$ is the unique $H^1(D)^2$ solution to the following boundary-value problem:

$$(3.19) \quad \begin{cases} -\operatorname{div}(A(h, \xi)e(p_{h,\xi})) = -2\chi(x)(u_{h,\xi} - u_T) & \text{in } D, \\ p_{h,\xi} = 0 & \text{on } \Gamma_D, \\ A(h, \xi)e(p_{h,\xi})n = 0 & \text{on } \partial D \setminus \overline{\Gamma_D}. \end{cases}$$

We first consider the optimization of the mechanism in the ideal situation where the Young's modulus is known perfectly, i.e. we solve (3.18) with $\xi = 0$, and thus $E(x, \xi) = E_0(x)$. The computation requires 2,000 iterations of our optimization algorithm, for a CPU time of about 3h 15 mn. The resulting design and the convergence history are reported on Fig. 12 (b) and (c).

Remark 3.3. In practice, we use small penalizations of the objective function $D(h, \xi)$ by the volume $\operatorname{Vol}(h)$ and the compliance $C(h, \xi)$ of the structure. The former helps removing unnecessary material islands disconnected from the main structure, while the latter enforces the connectedness of the structure, in spite of its trend to develop thin hinges, mimicking pointwise junctions, see e.g. [115] about this common practice in the design of compliant mechanisms.

We now consider a situation where the Young's modulus $E \equiv E(x, \xi)$ is unknown: its expression (3.16) stems from the Karhunen-Loève expansion of the covariance kernel $\operatorname{Cov}(x, y)$, after truncation at $k = 10$ modes. We thus solve a distributionally robust version of the optimization problem (3.18), assuming a Wasserstein ambiguity set $\mathcal{A}_W$ based on the nominal law $\mathbb{P} = \delta_0$, see Section 2.2. The problem at hand reads:

$$(3.20) \quad \min_{h \in \mathcal{U}_{\text{ad}}} \left(\sup_{Q \in \mathcal{A}_W} \int_{\Xi} D(h, \xi) \, dQ(\xi) \right).$$

The parameter ε is set to $\varepsilon = 0.01$. Invoking Proposition 2.1, (3.20) rewrites under the more convenient form:

$$\min_{\substack{h \in \mathcal{U}_{\text{ad}}, \\ \lambda \geq 0}} \mathcal{D}_W(h, \lambda), \text{ where } \mathcal{D}_W(h, \lambda) = \lambda m + \lambda \varepsilon \int_{\Xi} \log \left(\int_{\Xi} e^{\frac{D(h, \zeta) - c(\xi, \zeta)}{\lambda \varepsilon}} \, d\nu_{\xi}(\zeta) \right) \, d\mathbb{P}(\xi),$$

where the probability measure ν_{ξ} in (2.7) has variance $\sigma^2 = 1e-1$.

We solve this problem for several values of the radius m of the Wasserstein ball; the results are reported on Fig. 13; each computation requires 2,000 iterations of our numerical algorithm, for a CPU time of about 13h. The resulting designs are quite similar from one another; note however the change in the orientation of the hinges sustaining the jaws between the cases $m = 0.5$ and $m = 1$. Let us also emphasize that it is particularly difficult to achieve a “true” black-and-white design in this example: in spite of the penalization of the objective function by the volume hinted at in Remark 3.3, the designs eliminate intermediate densities at a very slow rate.

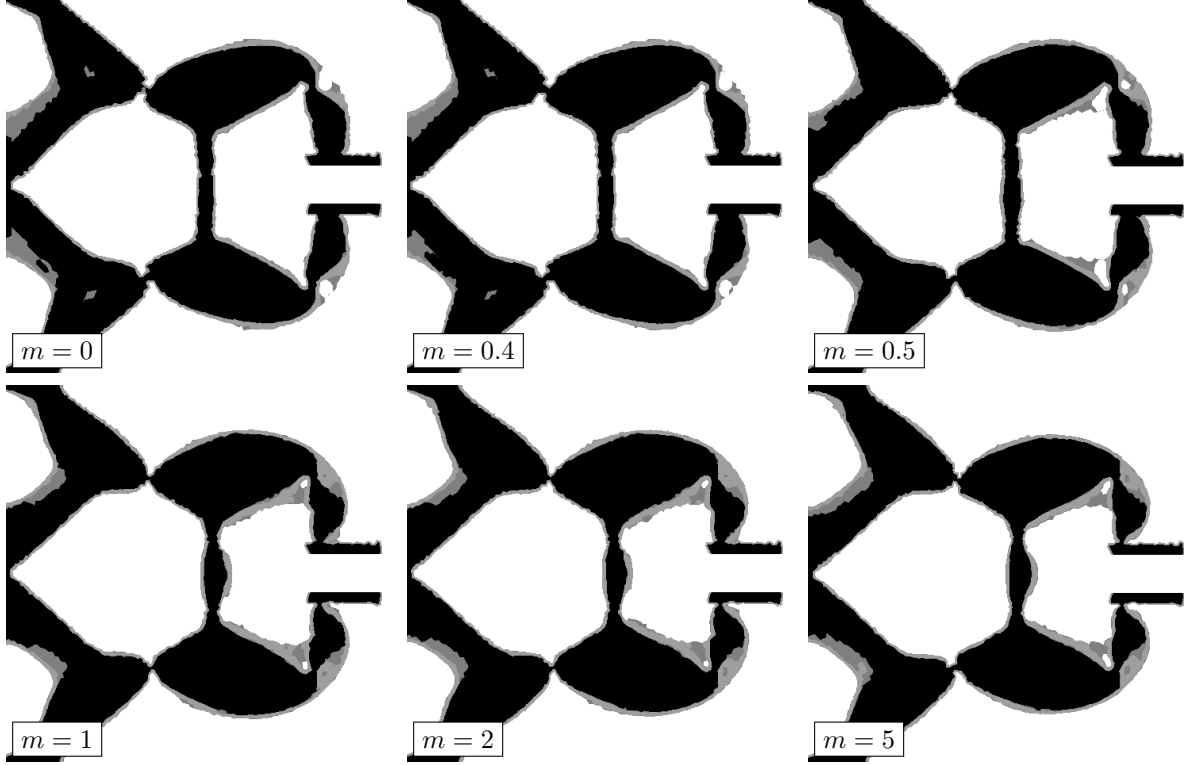


FIGURE 13. *Distributionally robust optimized designs for material uncertainties of the gripping mechanism considered in Section 3.5 associated to different values of the Wasserstein radius m .*

3.6. Reliability-based optimal design of a bridge using the conditional value at risk

This section illustrates the ideas of Section 2.4.2, regarding optimal design problems where the conditional value at risk of the cost function is used as a means to account for a failure probability. Before dealing with distributional robustness, we develop a little about this use of conditional value at risk, as it is relatively confidential in optimal design, to the best of our knowledge.

3.6.1. Deterministic optimal design of a bridge

The structures under consideration are 2d bridges, contained in a rectangular domain D with size 1×2 , meshed with approximately 8,500 vertices ($\approx 16,000$ triangles). They are clamped along the lower side Γ_D of ∂D , and a load is applied on their upper boundary Γ_N , as illustrated in Fig. 14 (a). In ideal conditions, this load equals $\xi^0 := (0, -1)$.

We first optimize the compliance $C(h, \xi^0)$ of the structure, defined by (3.6), under a volume constraint, in the ideal situation where the applied load is exactly ξ^0 , i.e. we solve:

$$(3.21) \quad \min_{h \in \mathcal{U}_{\text{ad}}} C(h, \xi^0) \text{ s.t. } \text{Vol}(h) = V_T, \text{ where } V_T = 0.245.$$

The resulting design h_{det}^* is depicted on Fig. 14 (b); its compliance equals $C(h_{\text{det}}^*, \xi^0) = 40.3892$.

3.6.2. Enforcement of a safety constraint via the conditional value at risk

We now assume that the load ξ applied on Γ_N is uncertain, but that its law $\mathbb{P} \in \mathcal{P}(\Xi)$ is known perfectly: it is (the restriction to a large ball $\Xi \subset \mathbb{R}^2$ of) the bivariate Gaussian law centered at ξ^0 , with variance σ^2 . We then wish to solve the following problem:

$$\min_{h \in \mathcal{U}_{\text{ad}}} \text{Vol}(h) \text{ s.t. } \mathbb{P}\left\{\xi \in \Xi \text{ s.t. } C(h, \xi) \geq C_T\right\} \leq 1 - \beta,$$

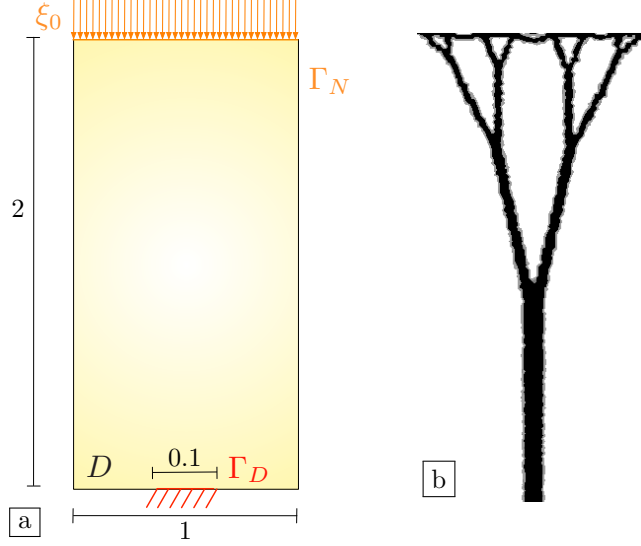


FIGURE 14. (a) Setting of the bridge problem considered in Section 3.6; (b) Optimized design h_{det}^* in ideal conditions.

where the threshold $C_T := 40$ is close to the optimal value $C(h_{\text{det}}^*, \xi^0)$ attained in the unperturbed minimization problem (3.21). According to the discussion in Section 2.4, we replace this problem by the following conservative version:

$$(3.22) \quad \min_{\substack{h \in \mathcal{U}_{\text{ad}}, \\ \alpha \in \mathbb{R}}} \text{Vol}(h) \quad \text{s.t.} \quad \alpha + \frac{1}{1-\beta} \int_{\xi \in \Xi} [C(h, \xi) - \alpha]_+ \, d\mathbb{P}(\xi) \leq C_T.$$

We solve this problem for several values of the variance σ^2 of the law $\mathbb{P}$ for ξ and the threshold β : the resulting designs are presented in Fig. 15, see also Table 3 for the associated values of the volume. In each situation, about 600 iterations of the optimization strategy described in Section 3.2 are used, for an approximate CPU cost of 2h 30 mn. Again, $N = 10$ samples are used for the numerical computation of each probabilistic integral featured in (3.22). Understandably enough, the closer β is to 1, the more stringent the failure constraint, and the larger the resulting volume. Loosely speaking, the most significant structural modifications occur in the lower vertical part of the domain, which tends to split into two sections joined by a reinforcement crossbar. We also note that it is more difficult for the design to converge towards clear black-and-white features (with the exact same numerical strategy), as intermediate densities seem to become more favorable.

$\sigma^2 \backslash \beta$	0.01	0.1	0.5	0.9	0.99
$1e-2.5$	0.365525	0.398984	0.523988	0.835165	0.840873
$1e-2$	0.438366	0.474774	0.553006	0.634271	0.641044

TABLE 3. Volume $\text{Vol}(h)$ of the optimized structures in the bridge example of Section 3.6.2.

3.6.3. Distributionally robust optimal design considering failure probabilities

We now turn to the situation where the law of the load parameter ξ is itself uncertain, which raises the need for a distributionally robust formulation of the optimal design problem (3.22). The nominal distribution is made of a single observation, associated to the ideal load $\xi^0 = (0, -1)$: $\mathbb{P} = \delta_{\xi^0}$. Assuming an ambiguity set

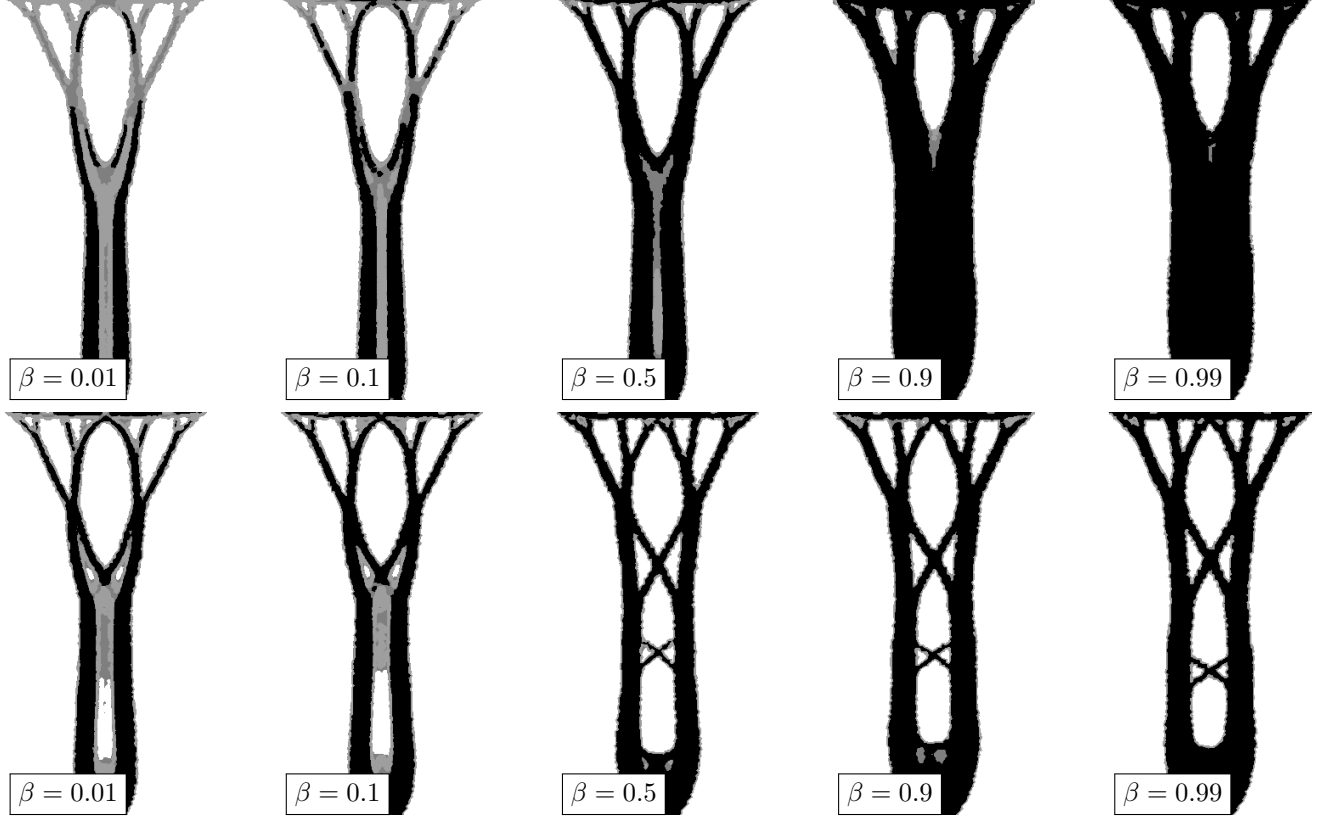


FIGURE 15. Optimized designs of the bridge under a safety constraint accounted for by the conditional value at risk, as considered in [Section 3.6.2](#); (Upper row) $\sigma^2 = 1e-2.5$; (Lower row) $\sigma^2 = 1e-2$.

of the Wasserstein type as in [Section 2.2.2](#), the distributionally robust version of the problem [\(3.22\)](#) reads:

$$\min_{\substack{h \in \mathcal{U}_{\text{ad}}, \\ \alpha \in \mathbb{R}}} \text{Vol}(h) \text{ s.t. } \sup_{\mathbb{Q} \in \mathcal{A}_W} \min_{\alpha \in \mathbb{R}} \left(\alpha + \frac{1}{1-\beta} \int_{\xi \in \Xi} [\mathcal{C}(h, \xi) - \alpha]_+ \, d\mathbb{Q}(\xi) \right) \leq C_T.$$

According to [Section 2.4](#), this rewrites:

$$\min_{\substack{h \in \mathcal{U}_{\text{ad}}, \\ \lambda \geq 0, \alpha \in \mathbb{R}}} \text{Vol}(h) \text{ s.t. } \mathcal{D}_C(h, \lambda, \alpha) \leq C_T, \text{ where}$$

$$\mathcal{D}_C(h, \lambda, \alpha) = \alpha + \frac{\lambda m}{1-\beta} + \frac{\lambda \varepsilon}{1-\beta} \log \left(\int_{\Xi} e^{\frac{[\mathcal{C}(h, \zeta) - \alpha]_+ - \lambda c(\xi^0, \zeta)}{\lambda \varepsilon}} \, d\nu_{\xi^0}(\zeta) \right).$$

This program is solved for several values of the threshold β , for two different values of the variance σ^2 of the reference coupling ν_{ξ^0} , and for the radius $m = 0.5$. The results are depicted on [Figs. 16](#) and [17](#). According to intuition, when perturbations become sufficiently large, the optimization algorithm gives privilege to the diagonal bars over the central vertical pillar of the structure. Also, as the variance σ^2 increases, the constraint becomes increasingly difficult to satisfy.

4. Distributionally Robust Geometric Shape Optimization

This section applies the generic distributionally robust formulations introduced in [Section 2](#) in the context of geometric shape optimization. After presenting the mechanical and mathematical settings in [Section 4.1](#), as well as the employed numerical methods for shape optimization in [Section 4.2](#), we conduct two numerical experiments in this context in [Sections 4.3](#) and [4.4](#).

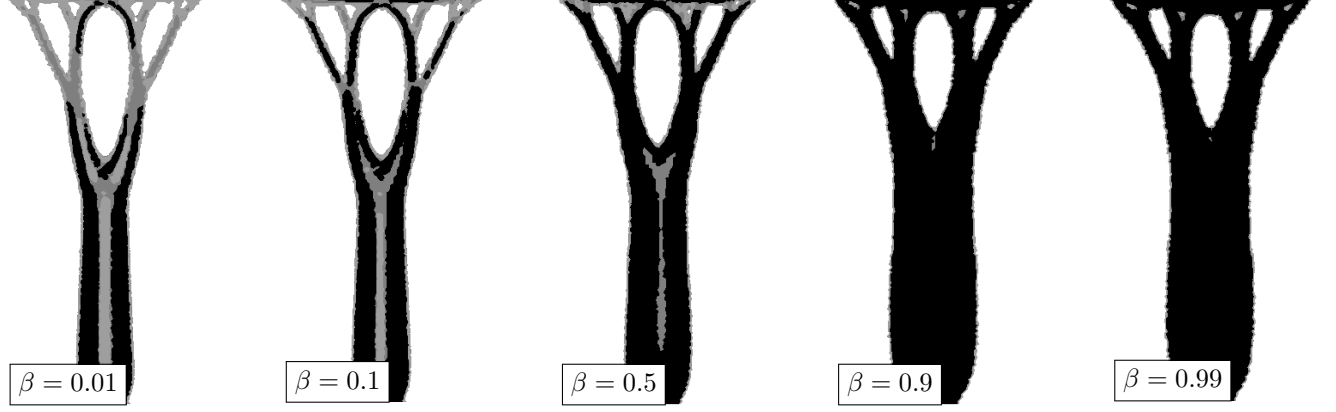


FIGURE 16. Distributionally robust designs of the bridge considered in [Section 3.6.3](#) for a Wasserstein radius $m = 0.5$, $\sigma^2 = 1e-3$ and various values of the parameter β .

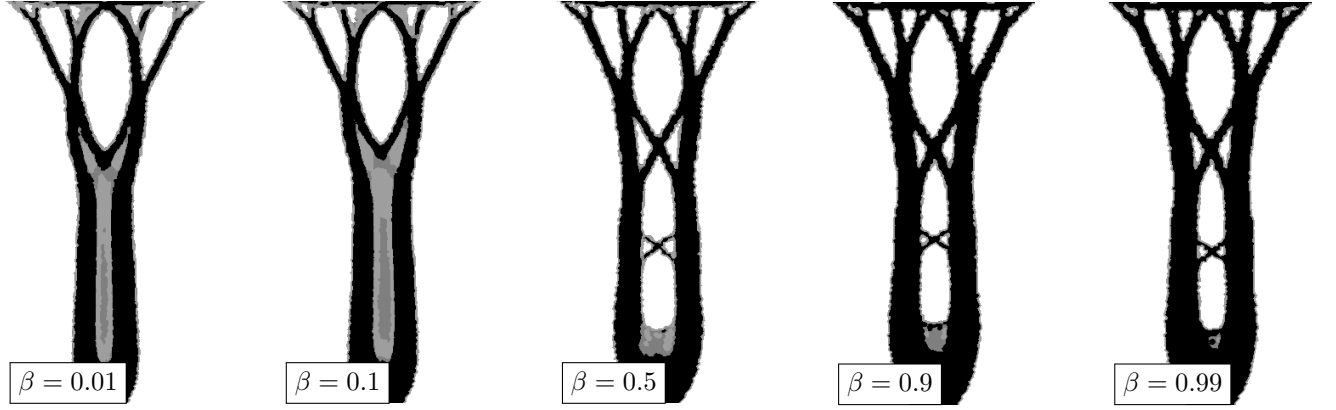


FIGURE 17. Distributionally robust designs of the bridge considered in [Section 3.6.3](#) for a Wasserstein radius $m = 0.5$, $\sigma^2 = 1e-2.5$ and various values of the parameter β .

4.1. Shape optimization of linear elastic structures

We first recall a few basic facts about the optimization of the shape of elastic structures.

4.1.1. Linear elastic structures

The designs considered henceforth are shapes, i.e. bounded Lipschitz domains $\Omega \subset \mathbb{R}^d$ ($d = 2$ or 3) representing elastic structures. The boundary $\partial\Omega$ of each shape Ω is decomposed into three disjoint regions:

$$\partial\Omega = \overline{\Gamma_D} \cup \overline{\Gamma_N} \cup \overline{\Gamma},$$

where:

- The shape is clamped on the region Γ_D ;
- Surface loads $g : \Gamma_N \rightarrow \mathbb{R}^d$ are applied on Γ_N ;
- The remaining part Γ is traction-free and it is the only region of $\partial\Omega$ which is subject to optimization.

Omitting body forces for simplicity, the elastic displacement of Ω is the unique solution $u \in H^1(\Omega)^d$ to the linear elasticity system:

$$(4.1) \quad \begin{cases} -\operatorname{div}(Ae(u)) = 0 & \text{in } \Omega, \\ u = 0 & \text{on } \Gamma_D, \\ Ae(u)n = g & \text{on } \Gamma_N, \\ Ae(u)n = 0 & \text{on } \Gamma. \end{cases}$$

Again, we have omitted body forces for simplicity. The isotropic and homogeneous properties of the constituent material of shapes are encoded in the Hooke's tensor A given by (3.1). In the sequel, we use various indices or arguments for u and related quantities of interest, in order to indicate their dependence on the relevant parameters of this model, notably the loads g and the shape Ω itself.

An archetypal shape optimization problem in this context reads:

$$(4.2) \quad \min_{\Omega} C(\Omega) \quad \text{s.t.} \quad \operatorname{Vol}(\Omega) = V_T,$$

where $C(\Omega)$ is the compliance of Ω , i.e.:

$$(4.3) \quad C(\Omega) = \int_{\Omega} Ae(u_{\Omega}) : e(u_{\Omega}) \, dx = \int_{\Gamma_N} g \cdot u_{\Omega} \, ds,$$

and $\operatorname{Vol}(\Omega) := \int_{\Omega} dx$ is the volume, which should match a target value V_T .

4.1.2. Shape sensitivity by the method of Hadamard

When it comes to study the sensitivity of a function with respect to the domain, we rely on the boundary variation method of Hadamard, about which we refer to e.g. [11, 58, 82, 103]. In brief, this method is based on variations of a reference shape $\Omega \subset \mathbb{R}^d$ of the form:

$$\Omega_{\theta} := (\operatorname{Id} + \theta)(\Omega), \quad \text{where } \theta \in W^{1,\infty}(\mathbb{R}^d, \mathbb{R}^d), \quad \|\theta\|_{W^{1,\infty}(\mathbb{R}^d, \mathbb{R}^d)} < 1,$$

i.e. Ω_{θ} is a version of Ω deformed by a “small” vector field θ , see Fig. 18.

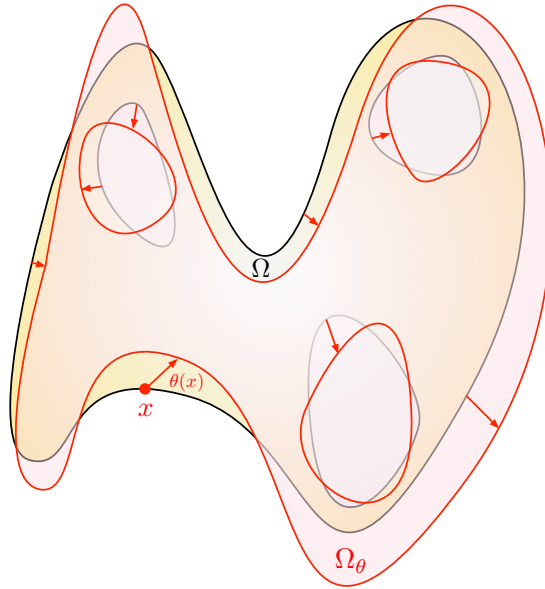


FIGURE 18. Variation $\Omega_{\theta} = (\operatorname{Id} + \theta)(\Omega)$ of a shape Ω by the method of Hadamard.

Definition 4.1. A functional $F(\Omega)$ is called shape differentiable at a particular domain Ω if the mapping

$$W^{1,\infty}(\mathbb{R}^d, \mathbb{R}^d) \ni \theta \mapsto F(\Omega_\theta) \in \mathbb{R}$$

is Fréchet differentiable at 0. Its derivative $F'(\Omega)(\theta)$ is called the shape derivative of F at Ω , and the following expansion holds true:

$$(4.4) \quad F(\Omega_\theta) = F(\Omega) + F'(\Omega)(\theta) + o(\theta), \text{ where } \frac{o(\theta)}{\|\theta\|_{W^{1,\infty}(\mathbb{R}^d, \mathbb{R}^d)}} \xrightarrow{\theta \rightarrow 0} 0.$$

It follows directly from this definition that a vector field θ satisfying $F'(\Omega)(\theta) < 0$ is a descent direction for $F(\Omega)$, i.e. for a small enough pseudo-time step $t > 0$, $F(\Omega_{t\theta}) < F(\Omega)$. More generally, in the context of a constrained optimization problem, of the form (4.3), the knowledge of the shape derivatives of the objective and constraint functions allows to infer a suitable “descent direction” via a constrained optimization method such as the Null-space algorithm used in this work [45]. The shape derivatives of the functions $F(\Omega)$ considered in this article often take the convenient form:

$$F'(\Omega)(\theta) = \int_{\partial\Omega} v_\Omega \theta \cdot n \, ds.$$

A descent direction for $F(\Omega)$ is immediately revealed from this structure, by taking $\theta = -v_\Omega n$ on $\partial\Omega$.

4.2. Numerical resolution of shape optimization problems by a level set based mesh evolution strategy

Our numerical resolution of shape optimization problems of the form (4.2) hinges on the level set based mesh evolution method introduced in [5, 6, 7], an open-source implementation of which is provided in [31]. Briefly, introducing a fixed computational domain $D \subset \mathbb{R}^d$, we rely on two complementary representations of each shape $\Omega \subset D$ arising during the optimization process:

- (Level set representation) As proposed in [87] – see also [10, 111] in the shape optimization context – Ω is represented implicitly, as the negative subdomain of a “level set” function $\phi : D \rightarrow \mathbb{R}$, i.e.

$$\begin{cases} \phi(x) < 0 & \text{if } x \in \Omega, \\ \phi(x) = 0 & \text{if } x \in \partial\Omega, \\ \phi(x) > 0 & \text{if } x \in D \setminus \bar{\Omega}. \end{cases}$$

In practice, such a function ϕ is discretized at the vertices of a simplicial mesh $\mathcal{T}$ of D (i.e. made of triangles in 2d, of tetrahedra in 3d), see Fig. 19 (a).

- (Meshed representation) The domain D is equipped with a conforming, high-quality simplicial mesh $\mathcal{T}$, made of two disjoint submeshes, $\mathcal{T}_{\text{int}}$ and $\mathcal{T}_{\text{ext}}$, which respectively represent Ω and its complement $D \setminus \bar{\Omega}$, see Fig. 19 (b).

Efficient numerical techniques make it possible to switch from one of these representations to the other:

- Given a meshed representation of a shape $\Omega \subset D$, one particular level set function for Ω is calculated as its signed distance function, thanks to the famous Fast Marching Algorithm [101]; this task relies on the open-source software `mshdist` from our previous work [32].
- Assuming a level set representation of $\Omega \subset D$, from the datum of the values of a level set function ϕ at the vertices of a mesh $\mathcal{T}$ of D , a new mesh $\tilde{\mathcal{T}}$ of D in which Ω is explicitly discretized is obtained thanks to a remeshing procedure, implemented in the open-source library `mmg`, see [15, 30].

Each operation of the shape optimization workflow is executed by using the most appropriate representation. In particular, “physical” boundary value problems posed on a shape $\Omega \subset D$ are solved by applying the Finite Element method on its meshed representation, which allows to use any numerical solver in a black-box manner; in our framework, we use the open-source platform `FreeFem` [57]. Conversely, the level set representation is particularly well-suited to track the evolution of shape $\Omega(t)$ over a time interval $(0, T)$, under the action of a prescribed velocity field $V(t, x)$. This evolution is indeed encoded in the following advection-like equation:

$$\frac{\partial \phi}{\partial t}(t, x) + V(t, x) \cdot \nabla \phi(t, x) = 0 \text{ for } t \in (0, T), x \in D,$$

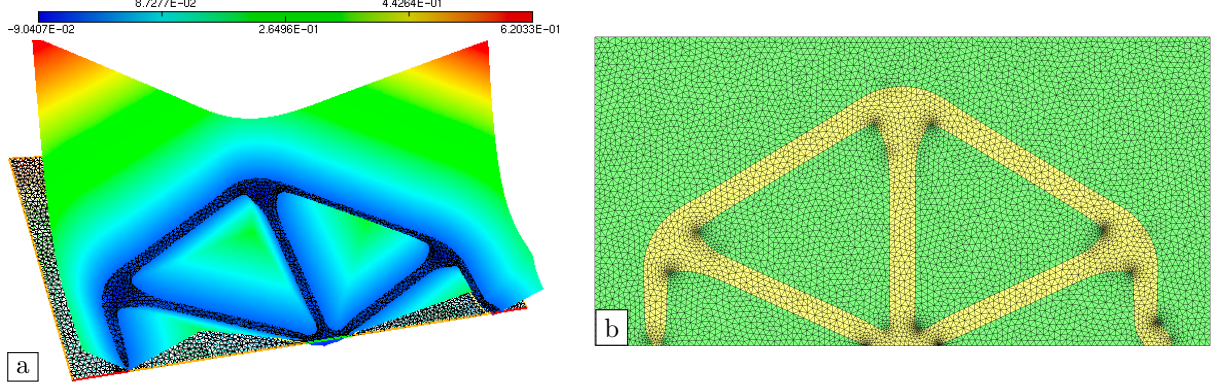


FIGURE 19. *Two complementary representations of a shape $\Omega \subset D$; (a) Graph of a level set function $\phi : D \rightarrow \mathbb{R}$; (b) Corresponding meshed representation.*

which can be solved e.g. by the method of characteristics [91]; we rely to this end on the open-source code `advection` from [21].

Summarizing, this strategy features a body-fitted mesh of the shape Ω at each stage of the workflow, allowing for precise mechanical analyses based on any finite element solver, while leaving the room for arbitrarily large deformations between successive optimization iterations.

4.3. Distributionally robust shape optimization of the stress within a 2d T-shaped beam under load uncertainties

Let the shape Ω account for a 2d T-shaped beam, enclosed in a box D with size 2×1 , see Fig. 20 (a). The beam is clamped on the upper region Γ_D of ∂D ; under ideal conditions, it is subjected to two identical vertical loads $\xi^0 = (0, -1)$, applied on two small regions at the left and right parts of ∂D , whose reunion is denoted by Γ_N . In the following, we denote by $u_{\Omega, \xi}$ the displacement of the beam, solution to (4.1) when the load ξ is applied.

We aim to minimize the total stress within Ω under a constraint on its volume, i.e. we consider the following shape optimization problem:

$$(4.5) \quad \min_{\Omega \subset D} S(\Omega, \xi^0) \quad \text{s.t.} \quad \text{Vol}(\Omega) = V_T.$$

Here, the stress function is defined by

$$S(\Omega, \xi) = \int_{\Omega} \chi(x) \|\sigma(u_{\Omega, \xi})\|^2 dx,$$

where the stress $\sigma(u)$ associated to a displacement field u is $\sigma(u) = Ae(u)$, and $\|\cdot\|$ denotes the usual Frobenius norm over matrices. The weight $\chi(x)$ equals 1 everywhere on D except on two small non-optimizable neighborhoods of the region Γ_N where it is set to 0 (these zones are depicted in blue in Fig. 20 (a)) Eventually, in (4.5), the target volume V_T is set to $V_T = 0.3$.

The shape derivative of the stress-based functional $S(\Omega, \xi)$ is given by the following proposition, which is taken from e.g. [9].

Proposition 4.1. *The shape derivative of the functional $S(\Omega, \xi)$ equals:*

$$S'(\Omega, \xi)(\theta) = \int_{\Omega} Ae(u_{\Omega, \xi}) : e(p_{\Omega, \xi}) \theta \cdot n ds,$$

where the adjoint state $p_{\Omega, \xi}$ is the unique solution in $H^1(\Omega)^2$ to the following boundary-value problem:

$$(4.6) \quad \begin{cases} -\text{div}(Ae(p_{\Omega, \xi})) = -2\text{div}(\chi(x)A\sigma(u_{\Omega, \xi})) & \text{in } \Omega, \\ p_{\Omega, \xi} = 0 & \text{on } \Gamma_D, \\ Ae(p_{\Omega, \xi})n = 2A\sigma(u_{\Omega, \xi})n & \text{on } \partial\Omega \setminus \overline{\Gamma_D}. \end{cases}$$

We first solve this ideal problem with the algorithm presented in [Section 4.2](#). About 300 iterations of our optimization algorithm are applied, corresponding to a computation of about 16 mn; the resulting shape Ω_{det}^* is depicted on [Fig. 20](#) (c). The corresponding value of the stress functional is $S(\Omega_{\text{det}}^*, \xi^0) = 1.99433$. The meshes of D produced at each iteration of the optimization process, made of submeshes for the structure Ω and its complement, contain about 6,500 vertices ($\approx 13,000$ triangles).

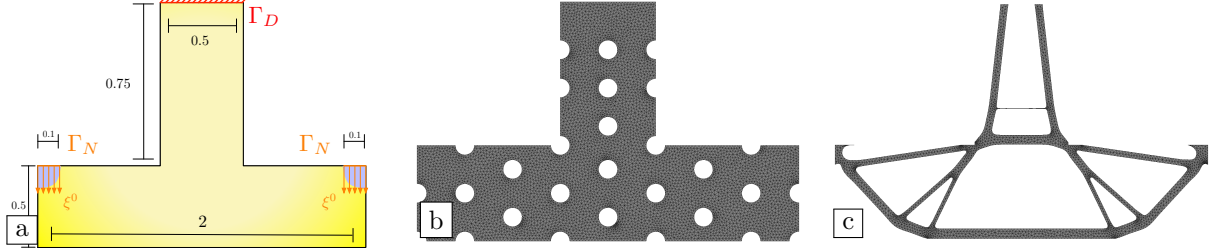


FIGURE 20. (a) Setting of the minimization of stress within a T-shaped beam considered in [Section 4.3](#); (b) Mesh of the initial shape; (c) Optimized shape Ω_{det}^* for the problem (4.5), assuming a perfect knowledge of the loads.

We now assume that the applied load ξ is uncertain; it belongs to a sufficiently large ball $\Xi \subset \mathbb{R}^2$, and its probability law is also uncertain – the only available information about the latter being that it is “close” to the nominal distribution $\mathbb{P} = \delta_{\xi^0} \in \mathcal{P}(\Xi)$ reconstructed from the single ideal load ξ^0 . We thus turn to a distributionally robust version of (4.5):

$$(4.7) \quad \min_{\Omega} \left(\sup_{\mathbb{Q} \in \mathcal{A}_W} \int_{\Xi} S(\Omega, \xi) d\mathbb{Q}(\xi) \right), \text{ s.t. } \text{Vol}(\Omega) = V_T,$$

where $\mathcal{A}_W$ is the Wasserstein ambiguity set defined in (2.9). Using the material in [Section 2.2](#), this problem rewrites:

$$\min_{\substack{\Omega, \\ \lambda \geq 0}} \mathcal{D}_W(\Omega, \lambda) \quad \text{s.t.} \quad \text{Vol}(\Omega) = V_T, \quad \text{where}$$

$$\mathcal{D}_W(\Omega, \lambda) = \lambda m + \lambda \varepsilon \int_{\Xi} \log \left(\int_{\Xi} \exp \left(\frac{S(\Omega, \zeta) - \lambda c(\xi, \zeta)}{\lambda \varepsilon} \right) d\nu_{\xi}(\zeta) \right) d\mathbb{P}(\xi).$$

Here, we recall that $\nu_{\xi}(\zeta) \in \mathcal{P}(\Xi) = \alpha_{\xi} e^{-\frac{|\xi - \zeta|^2}{2\sigma^2}} \mathbb{1}_{\Xi}(\zeta)$ is the probability measure featured in the definition of the reference coupling π_0 , see (2.7). The entropy regularization parameter ε is set to $\varepsilon = 0.01$. We solve this problem for several values of the Wasserstein radius m and the variance σ^2 . Each computation requires 300 iterations of our numerical strategy, for a CPU time of about 1h 20 mn. The numerical results are presented in [Fig. 21](#). The distributionally robust designs feature spread bars connected to the Dirichlet region Γ_D when compared to the result of [Fig. 20](#) (c). Likewise, the bars located at the center of the structure tend to adopt triangular or trapezoidal layouts, guaranteeing robustness under all perturbations.

The convergence histories of the computation are reported on [Fig. 22](#); the behavior of the objective function $\mathcal{D}_W$ is again typical of the use of the stochastic descent algorithm: a local minimum is attained after a very noisy path, due to the use of Monte-Carlo integration for the evaluation of probabilistic integrals.

The values of the stress functional for the various optimized shapes under ideal conditions (i.e. $\xi = \xi^0$) are reported on [Table 4](#). As expected, and like in the previous examples, these values deteriorate as the radius m increases. Moreover, we note that for small variance $\sigma^2 = 1e-3$ and $m = 0$, the stress within the distributionally robust structure is very close to that of the structure Ω_{det}^* .

4.4. Distributionally robust optimization of the shape of a 3d cantilever beam using a moment-based ambiguity set

This section deals with the optimization of a 3d cantilever beam, as depicted on [Fig. 23](#) (a). The structure under scrutiny is enclosed in a box D with size $2 \times 1 \times 1$; it is clamped on one face Γ_D of ∂D , and a load ξ

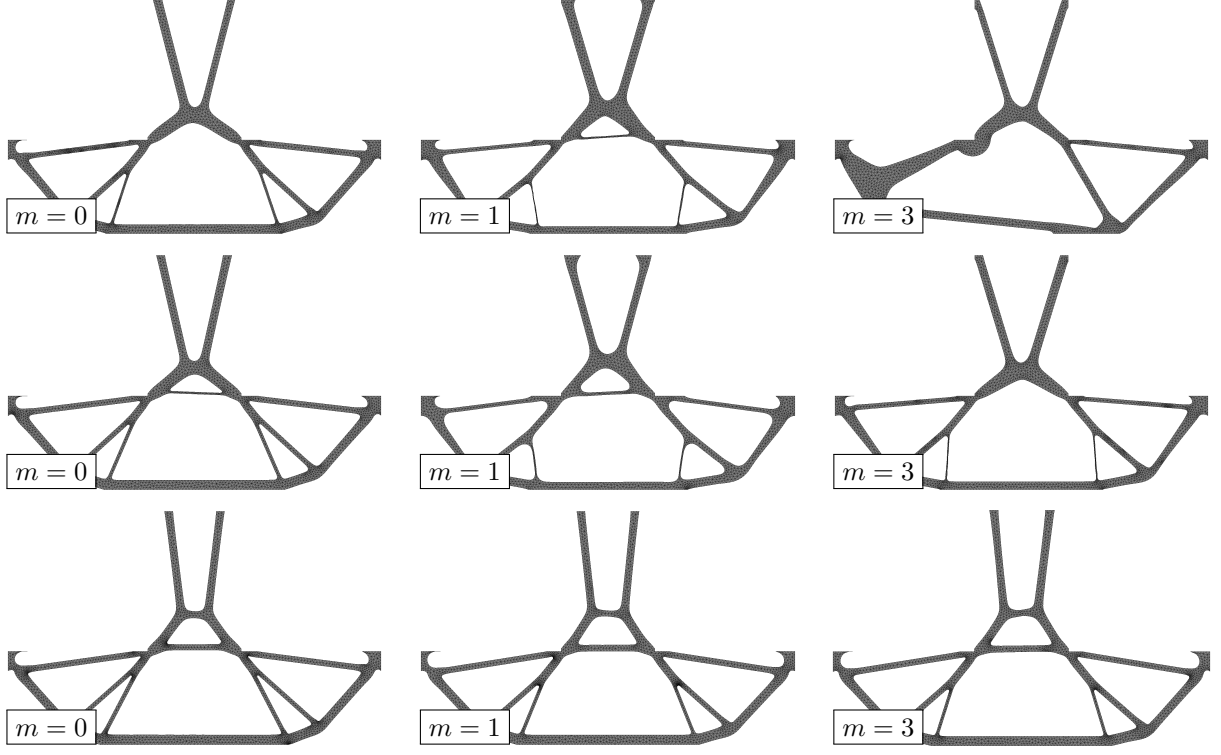


FIGURE 21. *Distributionally robust optimized designs obtained in the T-shaped beam example of Section 4.3, associated to various values of the parameters m and σ^2 ; (1st row) $\sigma^2 = 1e-0.5$; (2nd row) $\sigma^2 = 1e-1$; (3rd row) $\sigma^2 = 1e-3$.*

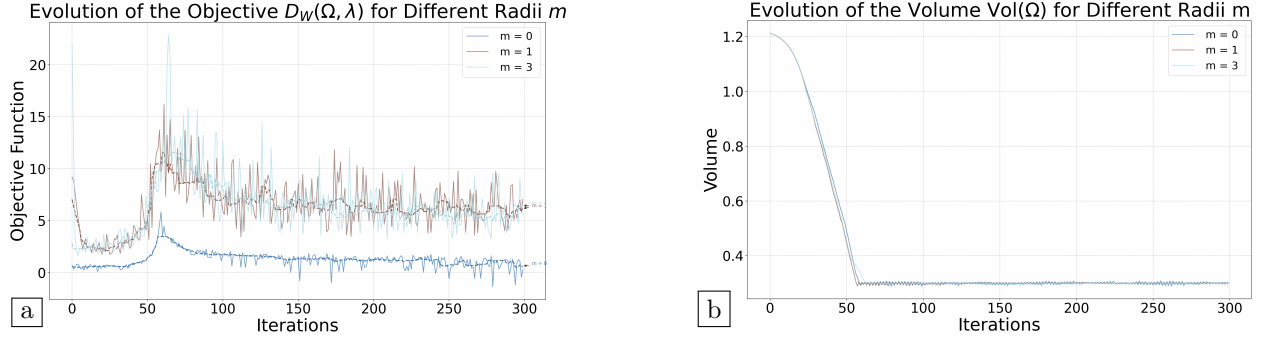


FIGURE 22. *Convergence histories of the objective function $D_W(\Omega, \lambda)$ and $\text{Vol}(\Omega)$ of the distributionally robust optimal result of the T-shaped beam of Fig. 21 for $\sigma^2 = 1e-1$.*

is applied on a small neighborhood Γ_N of the center of the opposite face. Like in the previous Section 4.3, we denote by $u_{\Omega, \xi}$ the solution to (4.1) when the load ξ is applied.

In an ideal setting, ξ is known perfectly, and it equals $\xi^0 = (0, 0, -1)$. We then consider the following problem:

$$(4.8) \quad \min_{\Omega} C(\Omega, \xi^0) \quad \text{s.t.} \quad \text{Vol}(\Omega) = V_T,$$

$\sigma^2 \backslash m$	0	1	3
$10^{-0.5}$	2.12165	2.32219	2.81476
10^{-1}	2.02971	2.27063	2.27391
10^{-3}	1.99752	2.01187	2.02397

TABLE 4. Values of the stress $S(\Omega, \xi^0)$ of the distributionally robust designs of the T-shaped beam considered in Section 4.3 when the ideal load ξ^0 is applied.

where $C(\Omega, \xi^0)$ is the compliance (4.3) of Ω and the volume target is set to $V_T = 0.45$. For convenience, let us recall the following formula for the shape derivative of the compliance, see e.g. [8, 10]:

$$C'(\Omega, \xi)(\theta) = - \int_{\Gamma} Ae(u_{\Omega, \xi}) : e(u_{\Omega, \xi}) \theta \cdot n \, ds.$$

The problem (4.8) is solved thanks to the numerical strategy presented in Section 4.2. Each mesh of D featured during the computation is composed of about 15,000 vertices ($\approx 80,000$ tetrahedra). The resulting shape Ω_{det}^* after 200 iterations ($\approx 2\text{h } 30\text{ mn}$ of computation) is depicted on Fig. 23 (b,c).

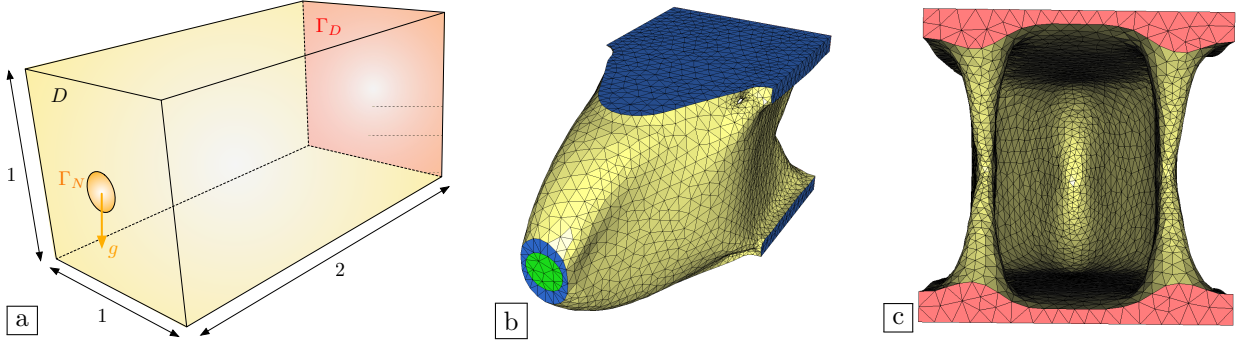


FIGURE 23. (a) Setting of the shape optimization example of a 3d cantilever beam considered in Section 4.4; (b) Front and (c) Back views of the optimized structure Ω_{det}^* in the ideal situation where the applied load ξ is perfectly known.

Now, we assume that the applied load ξ is uncertain: it belongs to a sufficiently large ball $\Xi \subset \mathbb{R}^3$ and its probability law is also unknown. The only information at hand about the latter is about its mean value $\mu^0 \in \mathbb{R}^3$ and covariance matrix $\Sigma^0 \in \mathbb{S}_+^3$:

$$\mu^0 = \xi^0 \quad \text{and} \quad \Sigma^0 = \sigma^2 \mathbf{I}, \quad \text{where } \sigma^2 = 0.01.$$

As proposed in Section 2.3, from these data, we now consider the following distributionally robust version of this problem:

$$\min_{\Omega} \left(\sup_{\mathbb{Q} \in \mathcal{A}_M} \int_{\Xi} C(\Omega, \xi) \, d\mathbb{Q}(\xi) - \varepsilon H(\mathbb{Q}) \right) \quad \text{s.t.} \quad \text{Vol}(\Omega) = V_T,$$

where $\mathcal{A}_M$ is the moment-based ambiguity-set defined in (2.13) and the entropy $H(\mathbb{Q})$ of a probability law $\mathbb{Q} \in \mathcal{P}(\Xi)$ equals (2.14). The penalty parameter ε is set to 0.001. This problem has the following equivalent expression:

$$\min_{\substack{\Omega, |\tau| \leq 1, \\ \lambda \geq 0, S \in \mathbb{S}_+^3}} \mathcal{D}_M(\Omega, \lambda, \tau, S), \quad \text{where} \\ \mathcal{D}_M(\Omega, \lambda, \tau, S) := \lambda m_1 - \lambda \tau \cdot \mu_0 + m_2 S : \Sigma_0 + \varepsilon \log \left(\int_{\Xi} \left(e^{\frac{C(\Omega, \xi) + \lambda \tau \cdot \xi - S : (\xi - \mu_0) \otimes (\xi - \mu_0)}{\varepsilon}} \right) d\mathbb{Q}_0(\xi) \right).$$

We solve this problem with the numerical strategy presented in [Section 4.2](#) for several values of the parameters m_1, m_2 . The optimized shapes obtained after 300 iterations (corresponding to approximately 3h 30 mn of computation) are depicted on [Figs. 24](#) and [25](#). The values of their compliance in ideal circumstances are reported on [Table 5](#).

	Ω_{det}^*	$\Omega_{m_1=0, m_2=1}^*$	$\Omega_{m_1=1, m_2=1}^*$	$\Omega_{m_1=2, m_2=1}^*$	$\Omega_{m_1=5, m_2=1}^*$	$\Omega_{m_1=0, m_2=2}^*$	$\Omega_{m_1=2, m_2=2}^*$	$\Omega_{m_1=5, m_2=5}^*$
$C(\Omega, \xi^0)$	0.0907753	0.0948306	0.0913598	0.108606	0.0979698	0.0903963	0.0923416	0.0978554

TABLE 5. *Values of the compliance $C(\Omega, \xi^0)$ of the distributionally robust designs of the 3d cantilever beam obtained in [Section 4.4](#) under ideal conditions.*

The optimized structures are not symmetric, which can be attributed to the unstructured nature of the mesh and the use of the stochastic gradient algorithm. Also note that, contrary to intuition, the values of the compliance of the obtained designs in ideal conditions are not strictly increasing functions of the parameters m_1, m_2 defining the ambiguity set $\mathcal{A}_M$. This phenomenon can be attributed to the fact that shape optimization problems of the form [\(4.8\)](#) generally have multiple local minima, so that the optimization path (and so the resulting optimized design) is generally very sensitive to the initial design, the mesh size, the optimization parameters, etc. In the present situation, this trend is even further exacerbated by the use of a stochastic descent algorithm. This intuition is confirmed by unreported experiments, that can be found in [\[93\]](#).

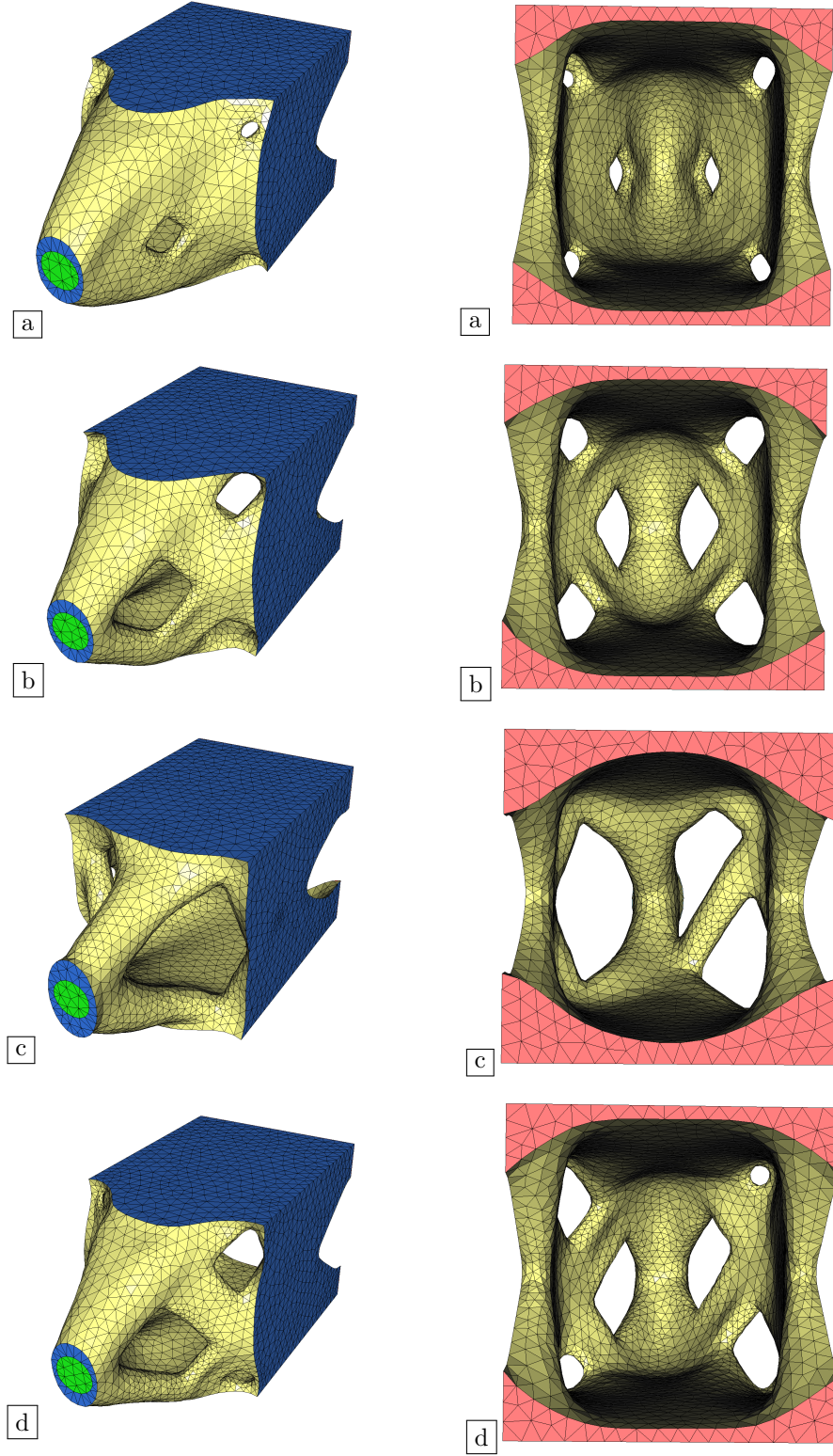


FIGURE 24. Optimized shapes of the 3d cantilever of [Section 4.4](#) under distributional uncertainties using a moment ambiguity set; front and back views for the parameters (a) $m_1 = 0$, $m_2 = 1$; (b) $m_1 = 1$, $m_2 = 1$; (c) $m_1 = 2$, $m_2 = 1$, and (d) $m_1 = 5$, $m_2 = 1$.

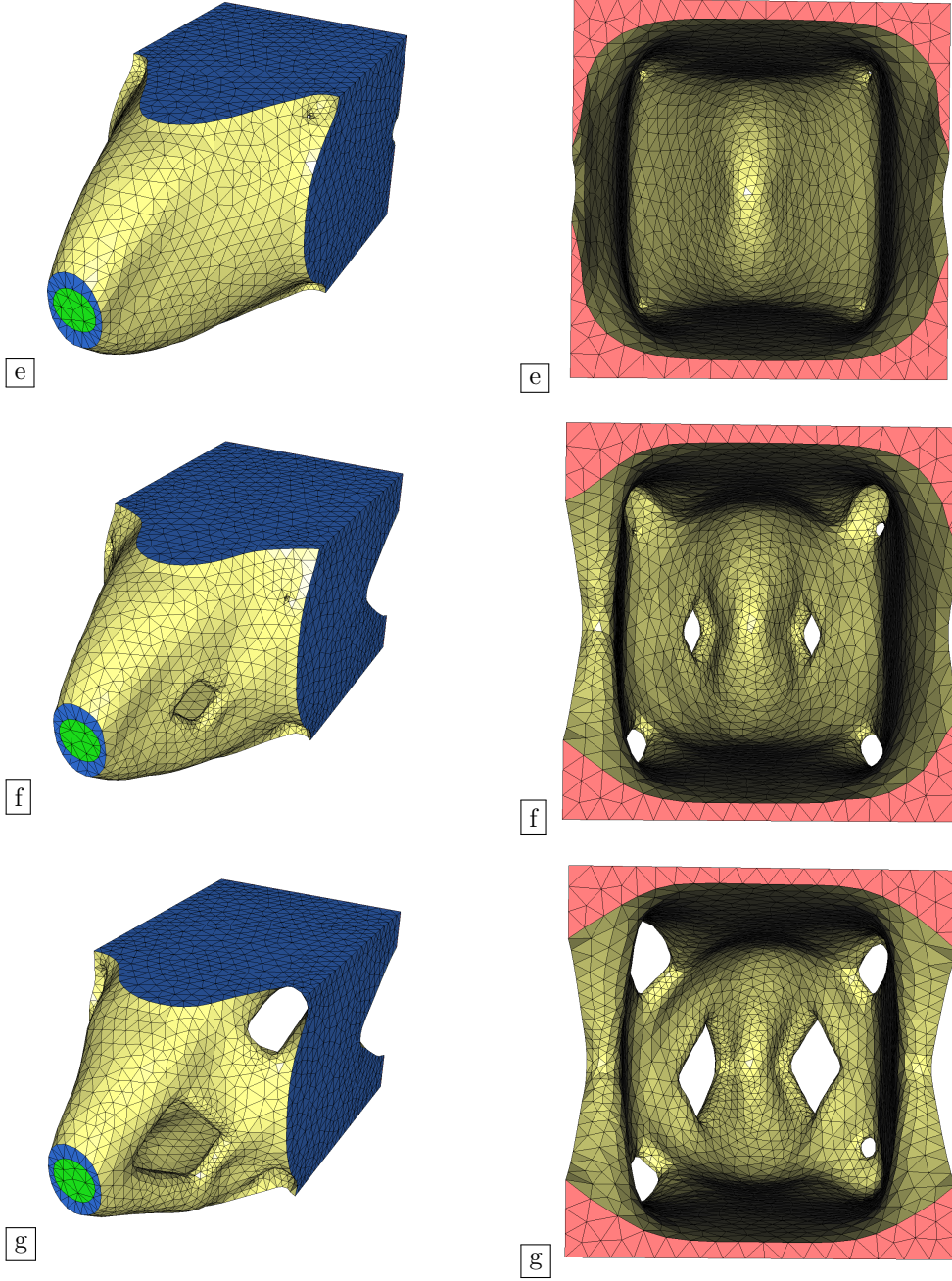


FIGURE 25. Optimized shapes of the 3d cantilever of [Section 4.4](#) under distributional uncertainties using a moment ambiguity set; front and back views for the parameters (e) $m_1 = 0$, $m_2 = 2$; (f) $m_1 = 2$, $m_2 = 1$, and (g) $m_1 = 5$, $m_2 = 5$.

5. Conclusion and Perspectives

This article was concerned with exploring the idea of distributional robustness in the context of shape and topology optimization: acknowledging that the physical parameters characterizing the behavior of real-life devices are inherently known with uncertainty, and that so is their probability law, we have proposed tractable robust formulations for a wide range of optimal design problems, that incorporate a degree of awareness about these sources of uncertainty. These formulations are not limited to a specific optimal design context and they can be used in both frameworks of density-based topology optimization and geometric shape optimization, in two and three space dimensions. As demonstrated by the numerical results, one key feature of distributionally robust optimal design formulations is their ability to anticipate scenarios “close”, but different from the observations used for the empirical reconstruction of the probability law of the uncertain parameters.

This work paves the way to research avenues of various kinds. From the theoretical viewpoint, we would like to justify rigorously that our tractable distributionally robust formulations, based on “small” penalizations by suitable entropy functionals, are consistent with their exact counterparts, as the penalty parameter tends to 0. From the numerical viewpoint, an obvious perspective concerns the choice of the optimization algorithm. The stochastic gradient indeed produces highly oscillatory trajectories, which can hinder convergence and degrade the quality of the final solution, particularly in high-variance settings. A different optimization algorithm, such as ADAM [65], could help to reduce the variance of descent direction estimates. Furthermore, the naive sampling strategy used in this work for evaluating probabilistic integrals could be improved, for instance by using variance reduction techniques such as importance sampling. From the modeling viewpoint, manifold physical situations suffer from uncertainty over physical parameters, and we wish to apply our techniques in e.g. thermal conduction, fluid mechanics or electromagnetism, to name a few. On a different note, the particular situation where the geometry of shapes itself is uncertain has been ignored in this work, in spite of its ubiquity and importance in applications. We believe that, with the help of the linearization techniques similar to those introduced in our former works [3, 4], it is possible to give a computationally affordable formulation of such problems. A foretaste of this important avenue for future research can be consulted in the Ph.D. manuscript [93].

Acknowledgements. The work of C.D, J.P. and B.T. is partially supported by the projects ANR-24-CE40-2216 STOIQUES and ANR-22-CE46-0006 StableProxies, financed by the French Agence Nationale de la Recherche (ANR). This research has benefitted from multiple discussions with F. Iutzeler (Institut de Mathématiques de Toulouse), whose expertise and insights are gratefully acknowledged.

APPENDIX A. Hint of Proof of **Proposition 2.1**

Let $A := \sup_{\mathbb{Q} \in \mathcal{A}_W} \int_{\Xi} f(\xi) d\mathbb{Q}(\xi)$ be the maximum value at the left-hand side of the relation to be established (2.11). We classically start by rewriting A with the help of a Lagrange multiplier for the constraint $\mathbb{Q} \in \mathcal{A}_W$; it holds:

$$A = \sup_{\mathbb{Q} \in \mathcal{P}(\Xi)} \inf_{\lambda \geq 0} \left(\int_{\Xi} f(\zeta) d\mathbb{Q}(\zeta) + \lambda (m - W_{\varepsilon}(\mathbb{P}, \mathbb{Q})) \right).$$

In fact, the inner infimum in this formula equals 0 when $\mathbb{Q}$ is such that $W_{\varepsilon}(\mathbb{P}, \mathbb{Q}) \leq m$ and $-\infty$ otherwise.

Now, formally interchanging the infimum and supremum in the above formula, and invoking the definition (2.8) of the regularized Wasserstein distance $W_{\varepsilon}(\cdot, \cdot)$, we obtain:

$$\begin{aligned} A &= \inf_{\lambda \geq 0} \left(\lambda m + \sup_{\mathbb{Q} \in \mathcal{P}(\Xi)} \sup_{\substack{\pi \in \mathcal{P}(\Xi \times \Xi) \\ \pi_1 = \mathbb{P}, \pi_2 = \mathbb{Q}}} \left(\int_{\Xi} f(\zeta) d\mathbb{Q}(\zeta) - \lambda \int_{\Xi \times \Xi} c(\xi, \zeta) d\pi(\xi, \zeta) - \lambda \varepsilon H(\pi) \right) \right) \\ &= \inf_{\lambda \geq 0} \left(\lambda m + \sup_{\substack{\pi \in \mathcal{P}(\Xi \times \Xi) \\ \pi_1 = \mathbb{P}}} \int_{\Xi \times \Xi} (f(\zeta) - \lambda c(\xi, \zeta)) d\pi(\xi, \zeta) - \lambda \varepsilon H(\pi) \right). \end{aligned}$$

We now proceed to reformulate the inner supremum, that we denote by B . The latter ranges over all probability couplings π whose first marginal equals $\mathbb{P}$. Recalling that the entropy of a coupling π blows up

when π is not absolutely continuous with respect to π_0 , this supremum can be recast over couplings of the form

$$(A.1) \quad \pi = \alpha(\xi, \zeta) \pi_0, \text{ where } \alpha \in L^1(\Xi \times \Xi; d\pi_0) \text{ satisfies } \int_{\Xi} \alpha(\xi, \zeta) d\nu_{\xi}(\zeta) = 1 \text{ for } \mathbb{P}\text{-a.e. } \xi \in \Xi.$$

Hence,

$$(A.2) \quad B = \sup_{\substack{\alpha \in L^1(\Xi \times \Xi; d\pi_0) \\ \int_{\Xi} \alpha(\xi, \zeta) d\nu_{\xi}(\zeta) = 1}} \int_{\Xi \times \Xi} \left(f(\zeta) - \lambda c(\xi, \zeta) - \lambda \varepsilon \log \alpha(\xi, \zeta) \right) \alpha(\xi, \zeta) d\pi_0(\xi, \zeta).$$

It turns out that this maximization problem, involving a strictly concave objective function, can be solved explicitly. Indeed, according to the first-order optimality conditions, there exists a function $\mu(\xi)$ such that its unique solution $\alpha(\xi, \zeta)$ satisfies:

$$f(\zeta) - \lambda c(\xi, \zeta) - \lambda \varepsilon \log \alpha(\xi, \zeta) - \lambda \varepsilon + \mu(\xi) = 0,$$

and so:

$$\alpha(\xi, \zeta) = e^{\frac{f(\zeta) - \lambda c(\xi, \zeta) - \lambda \varepsilon + \mu(\xi)}{\lambda \varepsilon}}.$$

Recalling the normalization (A.1), the function $\mu(\xi)$ is characterized by:

$$e^{\frac{\mu(\xi)}{\lambda \varepsilon}} = \left(\int_{\Xi} e^{\frac{f(\zeta) - \lambda c(\xi, \zeta) - \lambda \varepsilon}{\lambda \varepsilon}} d\nu_{\xi}(\zeta) \right)^{-1},$$

whence:

$$\alpha(\xi, \zeta) = \left(\int_{\Xi} e^{\frac{f(\zeta) - \lambda c(\xi, \zeta)}{\lambda \varepsilon}} d\nu_{\xi}(\zeta) \right)^{-1} e^{\frac{f(\zeta) - \lambda c(\xi, \zeta)}{\lambda \varepsilon}}.$$

Eventually, injecting this expression in (A.2), a simple calculation yields:

$$B = \lambda \varepsilon \int_{\Xi} \log \left(\int_{\Xi} e^{\frac{f(\zeta) - \lambda c(\xi, \zeta)}{\lambda \varepsilon}} d\nu_{\xi}(\zeta) \right) d\mathbb{P}(\xi),$$

and the desired formula (2.11) follows immediately. $\square$

APPENDIX B. Hint of Proof of Proposition 2.2

We rely on an argument similar to that employed in Appendix A. Let $A := \sup_{\mathbb{Q} \in \mathcal{A}_M} \left(\int_{\Xi} f(\xi) d\mathbb{Q}(\xi) - \varepsilon H(\mathbb{Q}) \right)$ be the worst-case of the (penalized) mean value of the function $f(\xi)$ when $\mathbb{Q}$ runs through the moment-based ambiguity set $\mathcal{A}_M \subset \mathcal{P}(\xi)$. By expressing this constraint with Lagrange multipliers, we obtain:

$$\begin{aligned} A = \sup_{\mathbb{Q} \in \mathcal{P}(\Xi)} \inf_{\substack{\lambda \geq 0, \\ S \in \mathbb{S}_+^k}} & \left(\int_{\Xi} f(\xi) d\mathbb{Q}(\xi) + \lambda \left(m_1 - \left| \mu_0 - \int_{\Xi} \xi d\mathbb{Q}(\xi) \right| \right) \right. \\ & \left. + S : \left(m_2 \Sigma_0 - \int_{\Xi} (\xi - \mu_0) \otimes (\xi - \mu_0) d\mathbb{Q}(\xi) \right) - \varepsilon H(\mathbb{Q}) \right). \end{aligned}$$

Thanks to the simple formula

$$\forall \xi \in \mathbb{R}^k, \quad |\xi| = \sup_{\substack{\tau \in \mathbb{R}^k, \\ |\tau| \leq 1}} \tau \cdot \xi,$$

this rewrites:

$$\begin{aligned} A = \sup_{\mathbb{Q} \in \mathcal{P}(\Xi)} \inf_{\substack{\lambda \geq 0, \\ |\tau| \leq 1, \\ S \in \mathbb{S}_+^k}} & \left(\int_{\Xi} f(\xi) d\mathbb{Q}(\xi) + \lambda m_1 - \lambda \tau \cdot \left(\mu_0 - \int_{\Xi} \xi d\mathbb{Q}(\xi) \right) \right. \\ & \left. + S : \left(m_2 \Sigma_0 - \int_{\Xi} (\xi - \mu_0) \otimes (\xi - \mu_0) d\mathbb{Q}(\xi) \right) - \varepsilon H(\mathbb{Q}) \right). \end{aligned}$$

A formal exchange of the infimum and supremum in the above formula then yields:

$$\begin{aligned}
A &= \inf_{\substack{\lambda \geq 0, \\ |\tau| \leq 1, \\ S \in \mathbb{S}_+^k}} (\lambda m_1 - \lambda \tau \cdot \mu_0 + m_2 S : \Sigma_0 + B), \text{ where} \\
B &:= \sup_{\mathbb{Q} \in \mathcal{P}(\Xi)} \left(\int_{\Xi} f(\xi) d\mathbb{Q}(\xi) + \lambda \tau \cdot \int_{\Xi} \xi d\mathbb{Q}(\xi) - S : \int_{\Xi} (\xi - \mu_0) \otimes (\xi - \mu_0) d\mathbb{Q}(\xi) - \varepsilon H(\mathbb{Q}) \right). \\
&= \sup_{\mathbb{Q} \text{ a.c. w.r.t. } \mathbb{Q}_0} \left(\int_{\Xi} f(\xi) d\mathbb{Q}(\xi) + \lambda \tau \cdot \int_{\Xi} \xi d\mathbb{Q}(\xi) - S : \int_{\Xi} (\xi - \mu_0) \otimes (\xi - \mu_0) d\mathbb{Q}(\xi) - \varepsilon H(\mathbb{Q}) \right),
\end{aligned}$$

where we have used the fact that $H(\mathbb{Q})$ takes infinite values when $\mathbb{Q}$ is not absolutely continuous with respect to $\mathbb{Q}_0$. We now proceed to calculate the quantity B :

$$(B.1) \quad B := \sup_{\substack{\alpha \in L^1(\Xi, d\mathbb{Q}_0), \\ \int_{\Xi} \alpha d\mathbb{Q}_0 = 1}} \int_{\Xi} \left(f(\xi) + \lambda \tau \cdot \xi - S : (\xi - \mu_0) \otimes (\xi - \mu_0) - \varepsilon \log \alpha(\xi) \right) \alpha(\xi) d\mathbb{Q}_0(\xi).$$

The necessary optimality condition for this strictly concave maximization problem yields the existence of a real value μ such that the maximizer α satisfies:

$$f(\xi) + \lambda \tau \cdot \xi - S : (\xi - \mu_0) \otimes (\xi - \mu_0) - \varepsilon \log \alpha(\xi) - \varepsilon + \mu = 0.$$

By rearranging this identity and using the normalization condition for α , we then obtain:

$$\alpha(\xi) = \frac{1}{\int_{\Xi} \left(e^{\frac{f(\xi) + \lambda \tau \cdot \xi - S : (\xi - \mu_0) \otimes (\xi - \mu_0)}{\varepsilon}} \right) d\mathbb{Q}_0(\xi)} e^{\frac{f(\xi) + \lambda \tau \cdot \xi - S : (\xi - \mu_0) \otimes (\xi - \mu_0)}{\varepsilon}}.$$

Now inserting this expression for α in (B.1), we eventually end up with:

$$B = \varepsilon \log \left(\int_{\Xi} \left(e^{\frac{f(\xi) + \lambda \tau \cdot \xi - S : (\xi - \mu_0) \otimes (\xi - \mu_0)}{\varepsilon}} \right) d\mathbb{Q}_0(\xi) \right),$$

which leads to the desired result (2.16). □

APPENDIX C. Hint of Proof of Proposition 2.3

For simplicity, we restrict ourselves to providing a sketch of the proof of this result in the case where the cumulative distribution function $\Psi(h, \cdot)$ of the law $\mathbb{P}$ is continuous on $\mathbb{R}$, i.e. there is no subset of Ξ with positive measure where the cost function $\mathcal{C}(h, \cdot)$ takes a given value $t \in \mathbb{R}$. Note however that the result holds true in the general case, see [96]. Let us set

$$F(h, \alpha) = \alpha + \frac{1}{1 - \beta} \int_{\Xi} [\mathcal{C}(h, \xi) - \alpha]_+ d\mathbb{P}(\xi).$$

A simple calculation shows that:

$$\begin{aligned}
F(h, \alpha) &= \alpha + \frac{1}{1 - \beta} \int_{\{\xi \in \Xi, \mathcal{C}(h, \xi) \geq \alpha\}} (\mathcal{C}(h, \xi) - \alpha) d\mathbb{P}(\xi) \\
&= \frac{1}{1 - \beta} \left(\int_{\{\xi \in \Xi, \mathcal{C}(h, \xi) \geq \alpha_{\beta}(h)\}} \alpha d\mathbb{P}(\xi) + \int_{\{\xi \in \Xi, \mathcal{C}(h, \xi) \geq \alpha\}} (\mathcal{C}(h, \xi) - \alpha) d\mathbb{P}(\xi) \right),
\end{aligned}$$

where we have used (2.18) to pass from the first line to the second one. Let us now make out two cases:

- If $\alpha_\beta(h) < \alpha$, a simple calculation yields:

$$\begin{aligned}
F(h, \alpha) &= \frac{1}{1-\beta} \left(\int_{\{\xi \in \Xi, \alpha_\beta(h) \leq \mathcal{C}(h, \xi) \leq \alpha\}} \alpha \, d\mathbb{P}(\xi) + \int_{\{\xi \in \Xi, \mathcal{C}(h, \xi) \geq \alpha\}} \mathcal{C}(h, \xi) \, d\mathbb{P}(\xi) \right) \\
&= \frac{1}{1-\beta} \int_{\{\xi \in \Xi, \mathcal{C}(h, \xi) \geq \alpha_\beta(h)\}} \max(\mathcal{C}(h, \xi), \alpha) \, d\mathbb{P}(\xi) \\
&\geq \frac{1}{1-\beta} \int_{\{\xi \in \Xi, \mathcal{C}(h, \xi) \geq \alpha_\beta(h)\}} \mathcal{C}(h, \xi) \, d\mathbb{P}(\xi) \\
&= \phi_\beta(h).
\end{aligned}$$

- If $\alpha < \alpha_\beta(h)$, it holds similarly:

$$\begin{aligned}
F(h, \alpha) &= \frac{1}{1-\beta} \left(- \int_{\{\xi \in \Xi, \alpha \leq \mathcal{C}(h, \xi) \leq \alpha_\beta(h)\}} \alpha \, d\mathbb{P}(\xi) + \int_{\{\xi \in \Xi, \mathcal{C}(h, \xi) \geq \alpha_\beta(h)\}} \mathcal{C}(h, \xi) \, d\mathbb{P}(\xi) \right. \\
&\quad \left. + \int_{\{\xi \in \Xi, \alpha \leq \mathcal{C}(h, \xi) \leq \alpha_\beta(h)\}} \mathcal{C}(h, \xi) \, d\mathbb{P}(\xi) \right) \\
&= \phi_\beta(h) + \frac{1}{1-\beta} \int_{\{\xi \in \Xi, \alpha \leq \mathcal{C}(h, \xi) \leq \alpha_\beta(h)\}} (\mathcal{C}(h, \xi) - \alpha) \, d\mathbb{P}(\xi) \\
&\geq \phi_\beta(h).
\end{aligned}$$

This proves that $F(h, \alpha) \leq \phi_\beta(h)$ for all $\alpha \neq \alpha_\beta(h)$, while the definition of $\alpha_\beta(h)$ immediately implies that $F(h, \alpha_\beta(h)) = \phi_\beta(h)$; this terminates the proof of the proposition. $\square$

REFERENCES

- [1] E. ACAR, G. BAYRAK, Y. JUNG, I. LEE, P. RAMU, AND S. S. RAVICHANDRAN, *Modeling, analysis, and optimization under uncertainties: a review*, Structural and Multidisciplinary Optimization, 64 (2021), pp. 2909–2945.
- [2] C. ACERBI AND D. TASCHE, *On the coherence of expected shortfall*, Journal of banking & finance, 26 (2002), pp. 1487–1503.
- [3] G. ALLAIRE AND C. DAPOGNY, *A linearized approach to worst-case design in parametric and geometric shape optimization*, Mathematical Models and Methods in Applied Sciences, 24 (2014), pp. 2199–2257.
- [4] ———, *A deterministic approximation method in shape optimization under random uncertainties*, SMAI Journal of computational mathematics, 1 (2015), pp. 83–143.
- [5] G. ALLAIRE, C. DAPOGNY, AND P. FREY, *Topology and geometry optimization of elastic structures by exact deformation of simplicial mesh*, Comptes Rendus Mathématique, 349 (2011), pp. 999–1003.
- [6] ———, *A mesh evolution algorithm based on the level set method for geometry and topology optimization*, Structural and Multidisciplinary Optimization, 48 (2013), pp. 711–715.
- [7] ———, *Shape optimization with a level set based mesh evolution method*, Computer Methods in Applied Mechanics and Engineering, 282 (2014), pp. 22–53.
- [8] G. ALLAIRE, C. DAPOGNY, AND F. JOUVE, *Shape and topology optimization*, in Geometric partial differential equations, part II, A. Bonito and R. Nochetto eds., Handbook of Numerical Analysis, 22 (2021), pp. 1–132.
- [9] G. ALLAIRE AND F. JOUVE, *Minimum stress optimal design with the level set method*, Engineering analysis with boundary elements, 32 (2008), pp. 909–918.
- [10] G. ALLAIRE, F. JOUVE, AND A.-M. TOADER, *Structural optimization using sensitivity analysis and a level-set method*, Journal of computational physics, 194 (2004), pp. 363–393.
- [11] G. ALLAIRE AND M. SCHOENAUER, *Conception optimale de structures*, vol. 58, Springer, 2007.
- [12] S. AMSTUTZ AND M. CILIGOT-TRAVAIN, *A notion of compliance robustness in topology optimization*, ESAIM: Control, Optimisation and Calculus of Variations, 22 (2016), pp. 64–87.
- [13] W. AZIZIAN, F. IUTZELER, AND J. MALICK, *Exact generalization guarantees for (regularized) wasserstein distributionally robust models*, Advances in Neural Information Processing Systems, 36 (2023), pp. 14584–14596.
- [14] ———, *Regularization for wasserstein distributionally robust optimization*, ESAIM: Control, Optimisation and Calculus of Variations, 29 (2023), p. 33.
- [15] G. BALARAC, F. BASILE, P. BÉNARD, F. BORDEU, J.-B. CHAPELIER, L. CIRROTTOLA, G. CAUMON, C. DAPOGNY, P. FREY, A. FROEHLI, ET AL., *Tetrahedral remeshing in the context of large-scale numerical simulation and high performance computing*, MathematicS In Action, 11 (2022), pp. 129–164.
- [16] M. P. BENDSOE AND O. SIGMUND, *Topology optimization: theory, methods, and applications*, Springer Science & Business Media, 2013.
- [17] P. BILLINGSLEY, *Probability and measure*, John Wiley & Sons, 2017.

- [18] P. BLANCHARD, D. J. HIGHAM, AND N. J. HIGHAM, *Accurately computing the log-sum-exp and softmax functions*, IMA Journal of Numerical Analysis, 41 (2021), pp. 2311–2330.
- [19] J. BLANCHET, J. LI, S. LIN, AND X. ZHANG, *Distributionally robust optimization and robust statistics*, arXiv preprint arXiv:2401.14655, (2024).
- [20] J. BLANCHET, K. MURTHY, AND V. A. NGUYEN, *Statistical analysis of wasserstein distributionally robust estimators*, in Tutorials in Operations Research: Emerging optimization methods and modeling techniques with applications, INFORMS, 2021, pp. 227–254.
- [21] C. BUI, C. DAPOGNY, AND P. FREY, *An accurate anisotropic adaptation method for solving the level set advection equation*, International Journal for Numerical Methods in Fluids, 70 (2012), pp. 899–922.
- [22] E. L. CARDOSO, G. DA SILVA, AND A. T. BECK, *Robust topology optimization of compliant mechanisms with uncertainties in output stiffness*, International Journal for Numerical Methods in Engineering, 119 (2019), pp. 532–547.
- [23] J. CÉA, *Conception optimale ou identification de formes, calcul rapide de la dérivée directionnelle de la fonction coût*, ESAIM: Mathematical Modelling and Numerical Analysis, 20 (1986), pp. 371–402.
- [24] A. CHAUDHURI, M. NORTON, AND B. KRAMER, *Risk-based design optimization via probability of failure, conditional value-at-risk, and buffered probability of failure*, in AIAA Scitech 2020 Forum, 2020, p. 2130.
- [25] R. CHEN, I. C. PASCHALIDIS, ET AL., *Distributionally robust learning*, Foundations and Trends® in Optimization, 4 (2020), pp. 1–243.
- [26] S. CHEN, W. CHEN, AND S. LEE, *Level set based robust shape and topology optimization under random field uncertainties*, Structural and Multidisciplinary Optimization, 41 (2010), pp. 507–524.
- [27] A. CHERKAEV AND E. CHERKAEVA, *Optimal design for uncertain loading condition*, in Homogenization: In Memory of Serguei Kozlov, World Scientific, 1999, pp. 193–213.
- [28] M. CUTURI, *Sinkhorn distances: Lightspeed computation of optimal transport*, Advances in neural information processing systems, 26 (2013).
- [29] C. DAPOGNY, *totuto a tentative implementation of density-based topology implementation in python*, <https://github.com/dapogny/totuto>, 2023.
- [30] C. DAPOGNY, C. DOBRZYNSKI, AND P. FREY, *Three-dimensional adaptive domain remeshing, implicit domain meshing, and applications to free and moving boundary problems*, Journal of computational physics, 262 (2014), pp. 358–378.
- [31] C. DAPOGNY AND F. FEPPON, *Shape optimization using a level set based mesh evolution method: an overview and tutorial*, Hal preprint: <https://hal.archives-ouvertes.fr/hal-03881641>, (2022).
- [32] C. DAPOGNY AND P. FREY, *Computation of the signed distance function to a discrete contour on adapted triangulation*, Calcolo, 49 (2012), pp. 193–219.
- [33] C. DAPOGNY, F. IUTZELER, A. MEDA, AND B. THIBERT, *Entropy-regularized wasserstein distributionally robust shape and topology optimization*, Structural and Multidisciplinary Optimization, 66 (2023), p. 42.
- [34] S. DE, J. HAMPTON, K. MAUTE, AND A. DOOSTAN, *Topology optimization under uncertainty using a stochastic gradient-based approach*, 2019.
- [35] S. DE, K. MAUTE, AND A. DOOSTAN, *Reliability-based topology optimization using stochastic gradients*, 2021.
- [36] ———, *Topology optimization under microscale uncertainty using stochastic gradients*, 2021.
- [37] F. DE GOURNAY, G. ALLAIRE, AND F. JOUVE, *Shape and topology optimization of the robust compliance via the level set method*, ESAIM: Control, Optimisation and Calculus of Variations, 14 (2008), pp. 43–70.
- [38] R. DEGHANNASIRI, X. QIAN, AND E. R. DOUGHERTY, *Intrinsically bayesian robust karhunen-loève compression*, Signal Processing, 144 (2018), pp. 311–322.
- [39] E. DELAGE AND Y. YE, *Distributionally robust optimization under moment uncertainty with application to data-driven problems*, Operations research, 58 (2010), pp. 595–612.
- [40] D. DRUSVYATSKIY AND L. XIAO, *Stochastic optimization with decision-dependent distributions*, Mathematics of Operations Research, 48 (2023), pp. 954–998.
- [41] M. EIGEL, J. NEUMANN, R. SCHNEIDER, AND S. WOLF, *Risk averse stochastic structural topology optimization*, Computer Methods in Applied Mechanics and Engineering, 334 (2018), pp. 470–482.
- [42] P. M. ESFAHANI AND D. KUHN, *Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations*, Mathematical Programming, 171 (2018), pp. 115–166.
- [43] F. FEPPON, *Null Space Optimizer*, <https://null-space-optimizer.readthedocs.io/en/latest/>, 2023.
- [44] ———, *Density-based topology optimization with the null space optimizer: a tutorial and a comparison*, Structural and Multidisciplinary Optimization, 67 (2024), pp. 1–34.
- [45] F. FEPPON, G. ALLAIRE, AND C. DAPOGNY, *Null space gradient flows for constrained optimization with applications to shape optimization*, ESAIM: Control, Optimisation and Calculus of Variations, 26 (2020), p. 90.
- [46] J. FEYDY, *Analyse de données géométriques, au delà des convolutions*, PhD thesis, Université Paris-Saclay, 2020.
- [47] N. FOURNIER AND A. GUILLIN, *On the rate of convergence in wasserstein distance of the empirical measure*, 2013.
- [48] M. GRIESHAMMER, L. PFLUG, M. STINGL, AND A. UIHLEIN, *The continuous stochastic gradient method: part i—convergence theory*, Computational Optimization and Applications, 87 (2024), pp. 935–976.
- [49] ———, *The continuous stochastic gradient method: part ii—application and numerics*, Computational Optimization and Applications, 87 (2024), pp. 977–1008.
- [50] M. GRIGORIU, *Simulation of stationary non-gaussian translation processes*, Journal of engineering mechanics, 124 (1998), pp. 121–126.

- [51] J. K. GUEST AND T. IGUSA, *Structural optimization under uncertain loads and nodal locations*, Computer Methods in Applied Mechanics and Engineering, 198 (2008), pp. 116–124.
- [52] X. GUO, W. BAI, W. ZHANG, AND X. GAO, *Confidence structural robust design and optimization under stiffness and load uncertainties*, Computer Methods in Applied Mechanics and Engineering, 198 (2009), pp. 3378–3399.
- [53] X. GUO, W. ZHANG, AND L. ZHANG, *Robust structural topology optimization considering boundary uncertainties*, Computer Methods in Applied Mechanics and Engineering, 253 (2013), pp. 356–368.
- [54] K. M. HAMDIA, H. GHASEMI, X. ZHUANG, AND T. RABCUZUK, *Multilevel monte carlo method for topology optimization of flexoelectric composites with uncertain material properties*, Engineering Analysis with Boundary Elements, 134 (2022), pp. 412–418.
- [55] G. A. HANASUSANTO, V. ROITCH, D. KUHN, AND W. WIESEMANN, *A distributionally robust perspective on uncertainty quantification and chance constrained programming*, Mathematical Programming, 151 (2015), pp. 35–62.
- [56] A. M. HASOFER, *An exact and invariant first order reliability format*, J. Eng. Mech. Div., Proc. ASCE, 100 (1974), pp. 111–121.
- [57] F. HECHT, *New development in freefem++*, Journal of numerical mathematics, 20 (2012), pp. 251–266.
- [58] A. HENROT AND M. PIERRE, *Shape Variation and Optimization*, EMS Tracts in Mathematics Vol. 28, 2018.
- [59] E. HOLMBERG, C.-J. THORE, AND A. KLARBRING, *Worst-case topology optimization of self-weight loaded structures using semi-definite programming*, Structural and Multidisciplinary Optimization, 52 (2015), pp. 915–928.
- [60] Z. HU AND L. J. HONG, *Kullback-leibler divergence constrained distributionally robust optimization*, Available at Optimization Online, 1 (2013), p. 9.
- [61] L. JOFRE AND A. DOOSTAN, *Rapid aerodynamic shape optimization under parametric and turbulence model uncertainty: A stochastic gradient approach*, 2021.
- [62] Y. KANNO, *Structural reliability under uncertainty in moments: distributionally-robust reliability-based design optimization*, Japan Journal of Industrial and Applied Mathematics, 39 (2021), p. 195–226.
- [63] M. G. KAPTEYN, K. E. WILLCOX, AND A. B. PHILPOTT, *Distributionally robust optimization for engineering design under uncertainty*, International Journal for Numerical Methods in Engineering, 120 (2019), pp. 835–859.
- [64] V. KESHAVARZZADEH, F. FERNANDEZ, AND D. A. TORTORELLI, *Topology optimization under uncertainty via non-intrusive polynomial chaos expansion*, Computer Methods in Applied Mechanics and Engineering, 318 (2017), pp. 120–147.
- [65] D. P. KINGMA AND J. BA, *Adam: A method for stochastic optimization*, 2017.
- [66] D. P. KOURI AND T. M. SUROWIEC, *Risk-averse pde-constrained optimization using the conditional value-at-risk*, SIAM Journal on Optimization, 26 (2016), pp. 365–396.
- [67] D. KUHN, P. M. ESFAHANI, V. A. NGUYEN, AND S. SHAFIEEZADEH-ABADEH, *Wasserstein distributionally robust optimization: Theory and applications in machine learning*, in Operations research & management science in the age of analytics, Informs, 2019, pp. 130–166.
- [68] D. KUHN, S. SHAFIEE, AND W. WIESEMANN, *Distributionally robust optimization*, 2024.
- [69] M. LANGELAAR, *Topology optimization for multi-axis machining*, Computer Methods in Applied Mechanics and Engineering, 351 (2019), pp. 226–252.
- [70] B. S. LAZAROV, M. SCHEVENELS, AND O. SIGMUND, *Topology optimization considering material and geometric uncertainties using stochastic collocation methods*, Structural and Multidisciplinary optimization, 46 (2012), pp. 597–612.
- [71] F. LIN, X. FANG, AND Z. GAO, *Distributionally robust optimization: A review on theory and applications*, Numerical Algebra, Control & Optimization, 12 (2022), p. 159.
- [72] J. LIU, A. T. GAYNOR, S. CHEN, Z. KANG, K. SURESH, A. TAKEZAWA, L. LI, J. KATO, J. TANG, C. C. WANG, ET AL., *Current and future trends in topology optimization for additive manufacturing*, Structural and Multidisciplinary Optimization, pp. 1–27.
- [73] Q. LIU, J. WU, X. XIAO, AND L. ZHANG, *A note on distributionally robust optimization under moment uncertainty*, Journal of Numerical Mathematics, 26 (2018), pp. 141–150.
- [74] M. LOËVE, *Probability theory*, Courier Dover Publications, 2017.
- [75] J. MARTÍNEZ-FRUTOS AND D. HERRERO-PÉREZ, *Large-scale robust topology optimization using multi-gpu systems*, Computer Methods in Applied Mechanics and Engineering, 311 (2016), pp. 393–414.
- [76] J. MARTÍNEZ-FRUTOS, D. HERRERO-PÉREZ, M. KESSLER, AND F. PERIAGO, *Robust shape optimization of continuous structures via the level set method*, Computer Methods in Applied Mechanics and Engineering, 305 (2016), pp. 271–291.
- [77] K. MAUTE, *Topology optimization under uncertainty*, in Topology optimization in structural and continuum mechanics, (2014), pp. 457–471.
- [78] Q. MERIGOT AND B. THIBERT, *Optimal transport: discretization and algorithms*, in Handbook of Numerical Analysis, vol. 22, Elsevier, 2021, pp. 133–212.
- [79] G. MICHAELIDIS, *Manufacturing constraints and multi-phase shape and topology optimization via a level-set method*, PhD thesis, 2014.
- [80] S. MOHAMED, M. ROSCA, M. FIGURNOV, AND A. MNIH, *Monte carlo gradient estimation in machine learning*, Journal of Machine Learning Research, 21 (2020), pp. 1–62.
- [81] E. F. MONTESUMA, F. M. N. MBOULA, AND A. SOULOUMIAC, *Recent advances in optimal transport for machine learning*, IEEE Transactions on Pattern Analysis and Machine Intelligence, (2024).
- [82] F. MURAT AND J. SIMON, *Sur le contrôle par un domaine géométrique*, Pré-publication du Laboratoire d’Analyse Numérique, (76015), (1976).

- [83] H. NAKAO, R. JIANG, AND S. SHEN, *Distributionally robust partially observable markov decision process with moment-based ambiguity*, SIAM Journal on Optimization, 31 (2021), pp. 461–488.
- [84] V. A. NGUYEN, D. KUHN, AND P. MOHAJERIN ESFAHANI, *Distributionally robust inverse covariance estimation: The wasserstein shrinkage estimator*, Operations research, 70 (2022), pp. 490–515.
- [85] J. NOCEDAL AND S. J. WRIGHT, *Numerical optimization 2nd*, Springer, 2006.
- [86] M. NUTZ, *Introduction to entropic optimal transport*, Lecture notes, Columbia University, (2021).
- [87] S. OSHER AND J. A. SETHIAN, *Fronts propagating with curvature-dependent speed: algorithms based on hamilton-jacobi formulations*, Journal of computational physics, 79 (1988), pp. 12–49.
- [88] G. PEYRÉ, M. CUTURI, ET AL., *Computational optimal transport: With applications to data science*, Foundations and Trends in Machine Learning, 11 (2019), pp. 355–607.
- [89] G. PFLUG AND D. WOZABAL, *Ambiguity in portfolio selection*, Quantitative Finance, 7 (2007), pp. 435–442.
- [90] L. PFLUG, M. STINGL, AND A. UIHLEIN, *A stochastic method of moving asymptotes for topology optimization under uncertainty*, arXiv preprint arXiv:2410.19428, (2024).
- [91] O. PIRONNEAU, *Finite element methods for fluids*, Wiley Chichester, 1989.
- [92] R.-E. PLESSIX, *A review of the adjoint-state method for computing the gradient of a functional with geophysical applications*, Geophysical Journal International, 167 (2006), pp. 495–503.
- [93] J. PRANDO, *Distributionally robust shape and topology optimization*, PhD thesis, Université Grenoble Alpes, 2025.
- [94] H. RAHIMIAN AND S. MEHROTRA, *Distributionally robust optimization: A review*, arXiv preprint arXiv:1908.05659, (2019).
- [95] ———, *Frameworks and results in distributionally robust optimization*, Open Journal of Mathematical Optimization, 3 (2022), pp. 1–85.
- [96] R. T. ROCKAFELLAR AND S. URYASEV, *Conditional value-at-risk for general loss distributions*, Journal of banking & finance, 26 (2002), pp. 1443–1471.
- [97] R. T. ROCKAFELLAR, S. URYASEV, ET AL., *Optimization of conditional value-at-risk*, Journal of risk, 2 (2000), pp. 21–42.
- [98] F. SANTAMBROGIO, *Optimal transport for applied mathematicians*, Birkhäuser, 2015.
- [99] C. SCHWAB AND R. A. TODOR, *Karhunen-loève approximation of random fields by generalized fast multipole methods*, Journal of Computational Physics, 217 (2006), pp. 100–122.
- [100] T. SÉJOURNÉ, J. FEYDY, F.-X. VIALARD, A. TROUVÉ, AND G. PEYRÉ, *Sinkhorn divergences for unbalanced optimal transport*, arXiv preprint arXiv:1910.12958, (2019).
- [101] J. A. SETHIAN, *Fast marching methods*, SIAM review, 41 (1999), pp. 199–235.
- [102] O. SIGMUND AND K. MAUTE, *Topology optimization approaches*, Structural and Multidisciplinary Optimization, 48 (2013), pp. 1031–1055.
- [103] J. SOKOLOWSKI AND J.-P. ZOLÉSIO, *Introduction to shape optimization*, Springer, 1992.
- [104] M. STAIB AND S. JEGELKA, *Distributionally robust optimization and generalization in kernel methods*, Advances in Neural Information Processing Systems, 32 (2019).
- [105] M. TOOTKABONI, A. ASADPOURE, AND J. K. GUEST, *Topology optimization of continuum structures under uncertainty—a polynomial chaos approach*, Computer Methods in Applied Mechanics and Engineering, 201 (2012), pp. 263–275.
- [106] A. UIHLEIN, O. SIGMUND, AND M. STINGL, *A 140 line matlab code for topology optimization problems with probabilistic parameters*, arXiv preprint arXiv:2505.10421, (2025).
- [107] C. VILLANI, *Optimal transport: old and new*, vol. 338, Springer, 2009.
- [108] F. VINCENT, W. AZIZIAN, F. IUTZELER, AND J. MALICK, *skwdro: a library for wasserstein distributionally robust machine learning*, arXiv preprint arXiv:2410.21231, (2024).
- [109] C. WANG, B. XU, Q. MENG, J. RONG, AND Y. ZHAO, *Topology optimization of cast parts considering parting surface position*, Advances in Engineering Software, 149 (2020), p. 102886.
- [110] J. WANG, R. GAO, AND Y. XIE, *Sinkhorn distributionally robust optimization*, arXiv preprint arXiv:2109.11926, (2021).
- [111] M. Y. WANG, X. WANG, AND D. GUO, *A level set method for structural topology optimization*, Computer methods in applied mechanics and engineering, 192 (2003), pp. 227–246.
- [112] M.-C. YUE, Y. RYCHENER, D. KUHN, AND V. A. NGUYEN, *A geometric unification of distributionally robust covariance estimators: Shrinking the spectrum by inflating the ambiguity set*, arXiv preprint arXiv:2405.20124, (2024).
- [113] Y.-G. ZHAO AND T. ONO, *A general procedure for first/second-order reliability method (form/sorm)*, Structural safety, 21 (1999), pp. 95–112.
- [114] M. ZHOU, R. FLEURY, Y.-K. SHYY, H. THOMAS, AND J. BRENNAN, *Progress in topology optimization with manufacturing constraints*, in 9th AIAA/ISSMO Symposium on multidisciplinary analysis and optimization, 2002, p. 5614.
- [115] B. ZHU, X. ZHANG, H. ZHANG, J. LIANG, H. ZANG, H. LI, AND R. WANG, *Design of compliant mechanisms using continuum topology optimization: A review*, Mechanism and Machine Theory, 143 (2020), p. 103622.