Entropy-regularized Wasserstein distributionally robust shape and topology optimization

Charles Dapogny, Franck Iutzeler, Andrea Meda, Boris Thibert

ENTROPY-REGULARIZED WASSERSTEIN DISTRIBUTIONALLY ROBUST SHAPE AND TOPOLOGY OPTIMIZATION

C. DAPOGNY[1], F. IUTZELER[1], A. MEDA[1] AND B. THIBERT [1]

[1] *Univ. Grenoble Alpes, CNRS, Grenoble INP*[1], *LJK, 38000 Grenoble, France.*

Abstract. This brief note aims to introduce the recent paradigm of distributional robustness in the field of shape and topology optimization. Acknowledging that the probability law of uncertain physical data is rarely known beyond a rough approximation constructed from observed samples, we optimize the worst-case value of the expected cost of a design when the probability law of the uncertainty is "close" to the estimated one up to a prescribed threshold. The "proximity" between probability laws is quantified by the Wasserstein distance, attached to optimal transport theory. The classical entropic regularization technique in this field combined with recent results from convex duality theory allow to reformulate the distributionally robust optimization problem in a way which is tractable for computations. Two numerical examples are presented, in the different settings of density-based topology optimization and geometric shape optimization. They exemplify the relevance and applicability of the proposed formulation regardless of the retained optimal design framework.

Contents

1. Introduction

In realistic situations, the physical behavior of a design h depends on one or several parameters, collectively denoted by $\xi \in \Xi$; for instance, when h is a mechanical structure, ξ may account for the loads applied on h, or the coefficients of the constituent elastic material. These parameters ξ affect, often dramatically, the physical response $u_{h,\xi}$ and the cost $\mathcal{C}(h,\xi)$ of h, which raises the need to incorporate a degree of awareness to uncertainties over their values in optimal design procedures [6, 13].

The treatment of uncertainties in optimal design usually fits in one of the following two frameworks. When no information is available about the uncertain parameters ξ, except for a maximum bound m on their amplitude $||\xi||$, worst-case design approaches consist in minimizing the worst-case scenario $h \mapsto \sup_{||\xi|| \leq m} \mathcal{C}(h,\xi)$. When the law $\mathbb{P}_{\text{true}}$ of ξ is known, probabilistic approaches are usually preferred: the expectation $h \mapsto \int_\Xi \mathcal{C}(h,\xi)\, \mathrm{d}\mathbb{P}_{\text{true}}(\xi)$ (or another risk measure) of the cost can be minimized, see e.g. [19]. Both approaches suffer from major drawbacks: while worst-case approaches are often deemed too "pessimistic", insofar as the robust design may show poor nominal performance for the sake of providing for an unlikely worst-case

[1]Institute of Engineering Univ. Grenoble Alpes

scenario, the law $\mathbb{P}_{\text{true}}$ of ξ featured in probabilistic approaches is often unknown, and in practice chosen in a heuristic manner.

Recently, the paradigm of distributionally robust optimization has emerged as an elegant means to overcome this conceptual shortcoming of probabilistic formulations, see [12, 17, 21]. Building on a nominal probability law $\mathbb{P}$ for ξ, which is for instance constructed from observed data, it minimizes the worst value $\sup_{\mathbb{Q}} \int_{\Xi} \mathcal{C}(h,\xi)\,\mathrm{d}\mathbb{Q}(\xi)$ of the expected cost when the law $\mathbb{Q}$ of ξ is "close" to $\mathbb{P}$ within a given tolerance m.

Hitherto, this idea has been considered mainly in academic settings; the purpose of this note is to show how it can be applied in the field of optimal design. More precisely, we adopt the viewpoint in [10, 15] where the notion of "closeness" between measures is quantified by the Wasserstein distance from optimal transport theory [14, 18]. We then take advantage of the key results from convex duality proved in [4, 20] to reformulate the entropy-regularized version of the distributionally robust optimal design problem in a manner which is amenable to computations.

In principle, this methodology can be implemented in any optimal design framework. For simplicity, the main ideas are presented in a formal and non technical way, in the context of a model density-based topology optimization problem, but we also propose a numerical example in the setting of (geometric) shape optimization. A longer article, containing full mathematical details and extensive numerical experiments, is currently in preparation.

2. Presentation of the distributionally robust optimal design problem

2.1. The deterministic compliance minimization problem

The considered designs are elastic structures contained in a fixed hold-all domain $D \subset \mathbb{R}^d$, that are clamped on a region $\Gamma_D \subset \partial D$ and subjected to traction loads on a disjoint subset $\Gamma_N \subset \partial D$. They are represented as density functions h on D, i.e.

$$h \in \mathcal{U}_{\text{ad}}, \quad \mathcal{U}_{\text{ad}} := L^\infty(D, [0,1]),$$

where $h(x)$ equals 0 (resp. 1) at points $x \in D$ surrounded by void (resp. by material) and $h(x) \in (0,1)$ accounts for a "grayscale" region made of a fictitious mixture of material and void. The material properties within D are encoded in the Hooke's tensor $A(h)$, which is related to the density function h via the so-called SIMP law:

$$A(h)(x) = \Big(\eta + (1-\eta)h(x)^p\Big)A, \quad x \in D, \tag{2.1}$$

where A is the Hooke's law of the reference elastic material, and $\eta \ll 1$ is a small parameter mimicking the presence of void, see e.g. [5] about this classical setting.

In our model situation, the uncertain parameters are the loads ξ applied on Γ_N. These are assumed to be constant and they belong to a "large enough" closed ball $\Xi = \overline{B(0,R)} \subset \mathbb{R}^d$, for some $R > 0$. The displacement of a design $h \in \mathcal{U}_{\text{ad}}$ in response to the loads $\xi \in \Xi$ is then the solution $u_{h,\xi} : D \to \mathbb{R}^d$ to the linear elasticity system:

$$\left\{ \begin{array}{ll} -\text{div}(A(h)e(u_{h,\xi})) = 0 & \text{in } D, \\ u_{h,\xi} = 0 & \text{on } \Gamma_D, \\ A(h)e(u_{h,\xi})n = \xi & \text{on } \Gamma_N, \\ A(h)e(u_{h,\xi})n = 0 & \text{on } \partial D \setminus (\overline{\Gamma_D} \cup \overline{\Gamma_N}), \end{array} \right. \tag{2.2}$$

where $e(u) := \frac{1}{2}(\nabla u + \nabla u^T)$ denotes the strain tensor associated to a displacement field $u : D \to \mathbb{R}^d$.

In this context, when the load $\xi \in \Xi$ is known exactly, the optimization problem of interest reads

$$\min_{h \in \mathcal{U}_{\text{ad}}} \mathcal{C}(h,\xi) \quad \text{s.t.} \quad \text{Vol}(h) = V_T. \tag{2.3}$$

Here, the cost $\mathcal{C}(h,\xi)$ of a design $h \in \mathcal{U}_{\text{ad}}$ submitted to the loads $\xi \in \Xi$ is the compliance, i.e.

$$\mathcal{C}(h,\xi) = \int_D A(h)e(u_{h,\xi}) : e(u_{h,\xi})\,\mathrm{d}x,$$

the volume functional is denoted by $\text{Vol}(h) = \int_D h\,\mathrm{d}x$, and V_T is a volume target.

2.2. **The distributionally robust compliance minimization problem**

We now turn to the more realistic situation where the loads ξ applied on Γ_N are governed by a probability law $\mathbb{Q}$ on Ξ which is unknown. Fortunately, in most practical situations, $\mathbb{Q}$ can be estimated by a nominal law $\mathbb{P}$, which is typically the empirical sum of a series of observations $\xi_i \in \Xi$, $i = 1, \dots, N$:

$$\mathbb{P} = \frac{1}{N} \sum_{i=1}^{N} \delta_{\xi_i}.$$

Let $\mathcal{P}(\Xi)$ be the space of probability measures on the compact set $\Xi \subset \mathbb{R}^d$, which we equip with the Wasserstein distance stemming from optimal transport theory, and defined by:

$$W(\mathbb{P}, \mathbb{Q}) = \inf \int_{\Xi \times \Xi} c(\xi, \zeta) \, \mathrm{d}\pi(\xi, \zeta), \tag{2.4}$$

where the infimum is taken over transport plans $\pi \in \mathcal{P}(\Xi \times \Xi)$ between $\mathbb{P}$ and $\mathbb{Q}$. These are probability measures on the product space $\Xi \times \Xi$ whose first and second marginals π_1 and π_2 coincide with $\mathbb{P}$ and $\mathbb{Q}$, respectively. Intuitively, the (quadratic) ground cost $c(\xi, \zeta) := |\xi - \zeta|^2$ on the set Ξ of loads measures the cost of "moving a unit of mass" from ξ to ζ, and $\pi(\xi, \zeta)$ encodes the "quantity of mass" transported from ξ to ζ. We refer e.g. to [18] about the properties of this distance, and to [16] for an overview of its use in applications.

With these definitions at hand, the distributionally robust counterpart of (2.3) is

$$\min_{h \in \mathcal{U}_{\mathrm{ad}}} J_{\mathrm{dr}}(h) \quad \text{s.t.} \quad \mathrm{Vol}(h) = V_T, \tag{2.5}$$

where $J_{\mathrm{dr}}(h)$ is the worst (maximum) value of the expected cost when the law $\mathbb{Q} \in \mathcal{P}(\Xi)$ of the uncertain parameter ξ is at distance less than a given threshold m from the nominal law $\mathbb{P}$:

$$J_{\mathrm{dr}}(h) = \sup_{\substack{\mathbb{Q} \in \mathcal{P}(\Xi) \\ W(\mathbb{P}, \mathbb{Q}) \leq m}} \int_{\Xi} \mathcal{C}(h, \xi) \, \mathrm{d}\mathbb{Q}(\xi). \tag{2.6}$$

3. Entropic regularization of the distributionally robust problem

The distributionally robust optimal design problem (2.5) is hard to tackle as is. To alleviate this issue, we consider the entropy-regularized version of the Wasserstein distance proposed in [7]:

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

where $\varepsilon > 0$ is a "small" smoothing parameter, the entropy $H(\pi)$ of an element $\pi \in \mathcal{P}(\Xi \times \Xi)$ is defined by

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

and $\pi_0 \in \mathcal{P}(\Xi \times \Xi)$ is a reference coupling. According to [4], a judicious choice about π_0, offering nice theoretical guarantees, is provided by the following formula:

$$\pi_0(\xi, \zeta) = \mathbb{P}(\xi) \mathrm{d}\nu_\xi(\zeta), \quad \text{with} \quad \mathrm{d}\nu_\xi(\zeta) := \alpha_\xi e^{-\frac{c(\xi, \zeta)}{2\sigma}} \mathbb{1}_\Xi(\zeta) \mathrm{d}\zeta \tag{3.1}$$

for some $\sigma > 0$ and a normalization factor α_ξ ensuring that $\mathrm{d}\nu_\xi$ is a probability distribution on Ξ. Precisely, the above definition means that

$$\text{For all continuous functions } \varphi : \Xi \times \Xi \to \mathbb{R}, \quad \int_{\Xi \times \Xi} \varphi(\xi, \zeta) \, \mathrm{d}\pi_0(\xi, \zeta) = \int_\Xi \left(\int_\Xi \varphi(\xi, \zeta) \, \mathrm{d}\nu_\xi(\zeta) \right) \mathrm{d}\mathbb{P}(\xi).$$

We are now in position to introduce the entropy-regularized version of the problem (2.5):

$$\min_{h \in \mathcal{U}_{\mathrm{ad}}} J_{\mathrm{dr},\varepsilon}(h) \quad \text{s.t.} \quad \mathrm{Vol}(h) = V_T, \text{ where } J_{\mathrm{dr},\varepsilon}(h) := \sup_{\substack{\mathbb{Q} \in \mathcal{P}(\Xi) \\ W_\varepsilon(\mathbb{P}, \mathbb{Q}) \leq m}} \int_\Xi \mathcal{C}(h, \xi) \, \mathrm{d}\mathbb{Q}(\xi). \tag{3.2}$$

This program is intricate at first glance, as it features nested maximization and minimization problems. Fortunately, the functional $J_{\mathrm{dr},\varepsilon}(h)$ admits a convenient dual reformulation as a minimum, as expressed by the following result from convex analysis, whose mild assumptions are omitted, and which is proved in a more general context in [4] (see also [20]).

Proposition 3.1. *Let Ξ be a convex and compact subset of $\mathbb{R}^d$, $f : \Xi \to \mathbb{R}$ be a continuous function and let $\mathbb{P} \in \mathcal{P}(\Xi)$ be a probability measure. For a sufficiently small value of σ, the following equality holds:*

$$\sup_{W_\varepsilon(\mathbb{P},\mathbb{Q})\leq m} \int_\Xi f(\xi)\,\mathrm{d}\mathbb{Q}(\xi) = \inf_{\lambda\geq 0} \left\{ \lambda m + \lambda\varepsilon \int_\Xi \log\left(\int_\Xi e^{\frac{f(\zeta)-\lambda c(\xi,\zeta)}{\lambda\varepsilon}} \mathrm{d}\nu_\xi(\zeta) \right) \mathrm{d}\mathbb{P}(\xi) \right\}.$$

Taking advantage of this result, the distributionally robust optimization problem (3.2) rewrites:

$$\min_{\substack{h\in\mathcal{U}_{\mathrm{ad}} \\ \lambda\geq 0}} \mathcal{D}(h,\lambda) \text{ s.t. } \mathrm{Vol}(h) = V_T, \tag{3.3}$$

$$\text{where } \mathcal{D}(h,\lambda) := \lambda m + \lambda\varepsilon \int_\Xi \log\left(\int_\Xi e^{\frac{\mathcal{C}(h,\zeta)-\lambda c(\xi,\zeta)}{\lambda\varepsilon}} \mathrm{d}\nu_\xi(\zeta) \right) \mathrm{d}\mathbb{P}(\xi). \tag{3.4}$$

This new version (3.3) boils down to a single minimization problem for the pair (h,λ). It can be solved by a standard constrained optimization algorithm, once the derivatives of the objective functional $\mathcal{D}(h,\lambda)$ with respect to both variables h and λ are obtained, which follows from a standard (albeit a little tedious) adjoint-based calculation.

Remark 3.1. *A similar duality result to that of Proposition 3.1 actually holds when the regularized quantity $W_\varepsilon(\mathbb{P},\mathbb{Q})$ is replaced by the true Wasserstein distance $W(\mathbb{P},\mathbb{Q})$, leading to a reformulation of the distributionally robust problem (2.5) of the form (3.3). The latter is however more difficult to handle from the numerical viewpoint since it involves the supremum $\sup_{\zeta\in\Xi}\left(\mathcal{C}(h,\zeta) - \lambda c(\xi,\zeta)\right)$ in place of the "smooth" log-sum-exp approximation $\varepsilon \log\left(\int_\Xi e^{\frac{\mathcal{C}(h,\zeta)-\lambda c(\xi,\zeta)}{\lambda\varepsilon}} \mathrm{d}\nu_\xi(\zeta)\right)$ featured in (3.4), see [10].*

4. Numerical results

4.1. Topology optimization of a 2d bridge

Our first numerical example unfolds in the density-based context of Section 2.1, and deals with the topology optimization of a 2d bridge. The considered designs h are contained in a box D with size 1×1; they are clamped on the bottom side Γ_D of ∂D and subjected to a constant load ξ applied on the whole upper side Γ_N, see Fig. 1 (top, left) for details. The nominal probability law $\mathbb{P}$ for ξ is constructed from one single observation ξ_1, corresponding to a unit vertical load:

$$\mathbb{P} = \delta_{\xi_1}, \quad \xi_1 = (0,-1).$$

The entropic regularization coefficient ε is $1e^{-2}$ and the parameter σ appearing in the reference coupling π_0 in (3.1) equals $1e^{-3}$.

The distributionally robust topology optimization problem (3.3) is solved for several values of the tolerance parameter m and the target volume $V_T = 0.2$. The optimized designs are represented on Fig. 1, and the histories of the computation are displayed on Fig. 2. Understandably enough, the optimized designs develop thin branches to cope with larger loads, with horizontal components, and their nominal performance $\mathcal{C}(h,\xi_1)$ gets increasingly bad as m grows, see Table 1.

Value of m	0	0.25	0.52	0.6	0.9	1
Nominal compliance	13.9902	17.3063	19.2063	19.6829	24.3765	30.2474

Table 1. Values of the nominal cost for the optimized bridges of Section 4.1.

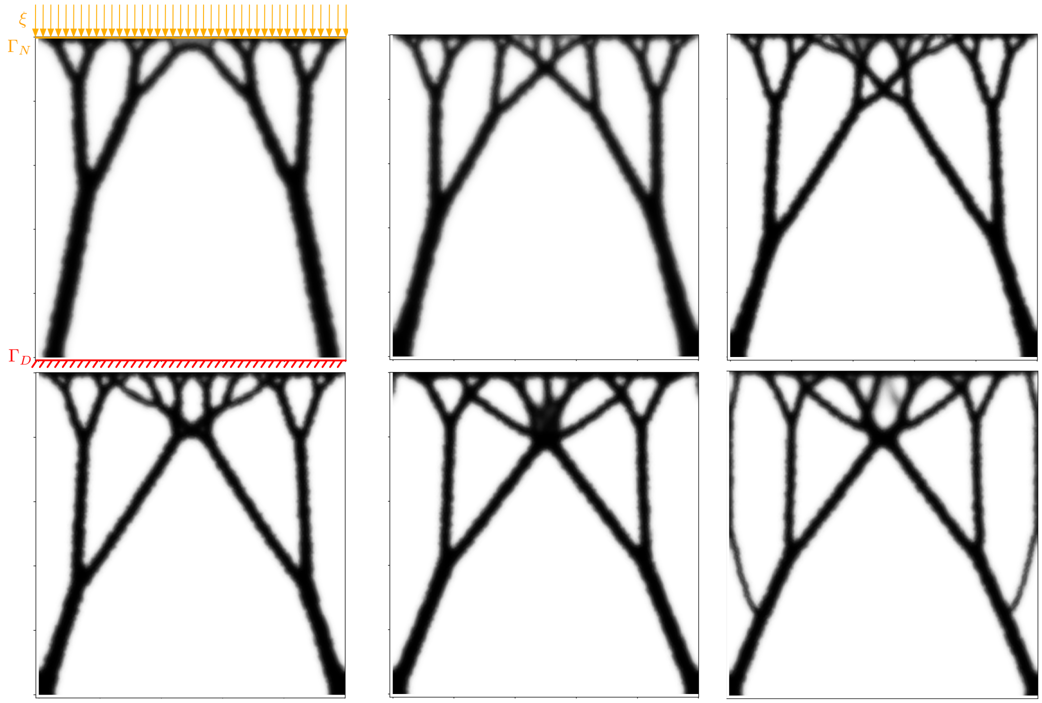

FIGURE 1. *(From left to right, top to bottom) Optimized density in the bridge topology optimization example of Section 4.1 for* $m = 0$ *(with details of the test-case) and* $m = 0.25$, 0.52, 0.6, 0.9, 1.

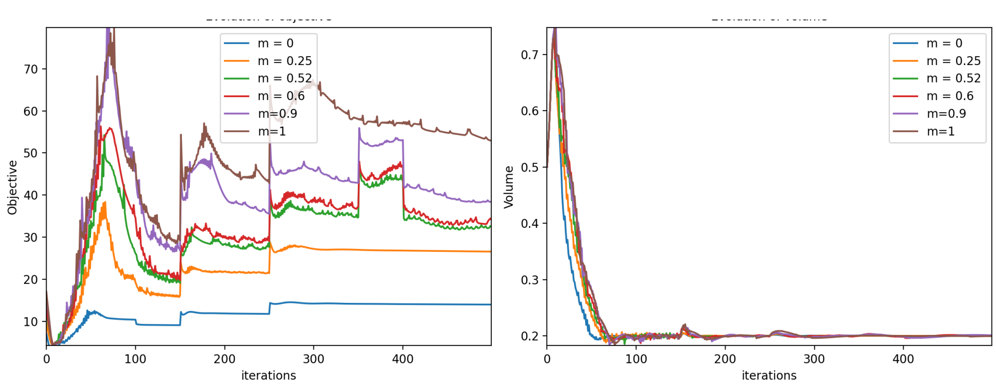

FIGURE 2. *Convergence histories for the experiments conducted with the bridge topology optimization example of Section 4.1; the large jumps in the values of the objective function at particular iterations are due to an increase in the parameter* p *of the material law* (2.1).

4.2. **Geometric optimization of a 2d cantilever**

Our second example is about the optimization of a 2d cantilever beam, and it is treated from the geometric shape optimization viewpoint [3, 2]: the considered designs are domains Ω, contained in the fixed computational domain $D = [0, 2] \times [0, 1]$. They are clamped on their left-hand side $\Gamma_D \subset \partial D$ and a constant load $\xi \in \Xi$ is applied on a small region Γ_N at the middle of their right-hand side, see Fig. 3 for the details. We rely on the mesh evolution method from our previous work [1] to track the evolution of the mesh of the optimized shape.

Again, the nominal law $\mathbb{P}$ for the load is constructed from only one sample ξ_1:

$$\mathbb{P} = \delta_{\xi_1}, \quad \xi_1 = (-1, 0).$$

The parameters ε and σ equal respectively $1e^{-2}$ and $1e^{-3}$.

We solve several instances of the distributionally robust problem (3.3) for various values of m and the volume target $V_T = 0.6$ with the help of the constrained optimization algorithm introduced in [9]; the results are reported on Fig. 3, see Fig. 4 for the histories of the computations. Again, the nominal performance of the designs tends to deteriorate when the size of the parameter m increases, see Table 2.

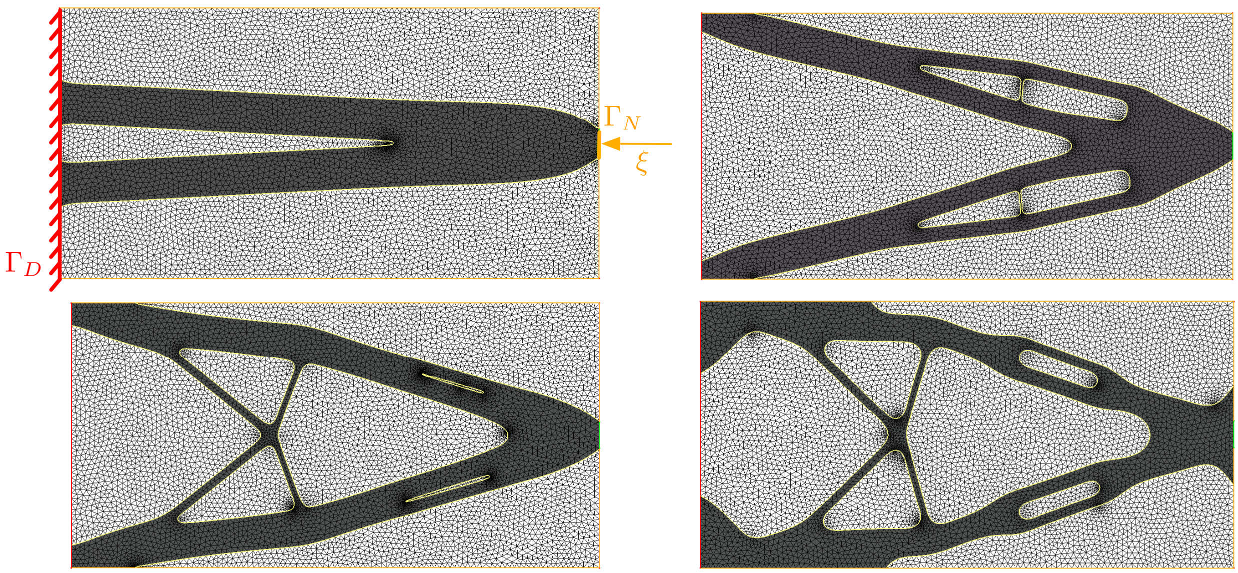

FIGURE 3. *(From left to right, top to bottom) Optimized shape the cantilever shape optimization example of Section 4.2 for* $m = 0$ *(with details of the test-case), optimized shapes for* $m = 1,\ 1.5,\ 2$.

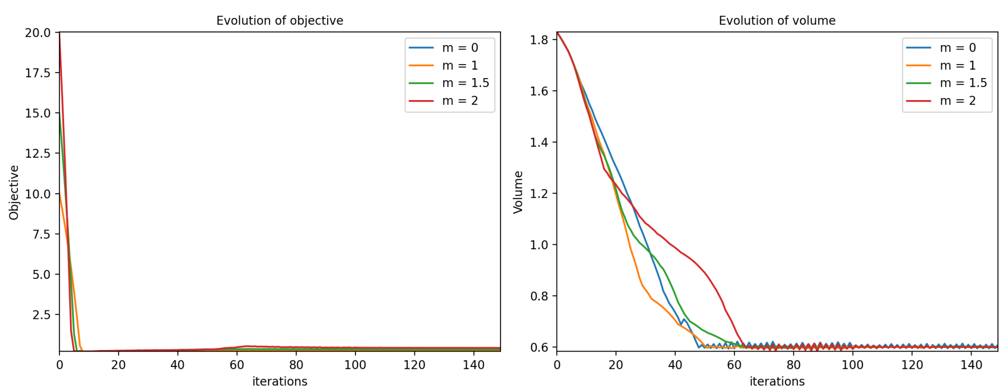

FIGURE 4. *Convergence histories for the experiments conducted with the cantilever shape optimization example of Section 4.2.*

Value of m	0	1	1.5	2
Nominal compliance	0.0646956	0.0746958	0.08138	0.0963393

TABLE 2. Values of the nominal cost for the optimized cantilever beams of Section 4.2.

Conflict of interest. On behalf of all authors, the corresponding author states that there is no conflict of interest.

Replication of results. The numerical example of Section 4.1 is tackled with the open-source finite element environment `FreeFem` [11], and the precise source code used for the resolution is available on demand. The treatment of the numerical example of Section 4.2 relies on minor adaptations to the open-source, educational implementation supplied with the article [8].

Acknowledgements. The work of C.D. was partially supported by the project ANR-18-CE40-0013 SHAPO financed by the French Agence Nationale de la Recherche (ANR).

References

[1] G. Allaire, C. Dapogny, and P. Frey, *Shape optimization with a level set based mesh evolution method*, Computer Methods in Applied Mechanics and Engineering, 282 (2014), pp. 22–53.

[2] 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, vol. 22, (2021), pp. 1–132.

[3] G. Allaire and M. Schoenauer, *Conception optimale de structures*, vol. 58, Springer, 2007.

[4] W. Azizian, F. Iutzeler, and J. Malick, *Regularization for wasserstein distributionally robust optimization*, arXiv preprint arXiv:2205.08826, (2022).

[5] M. P. Bendsoe and O. Sigmund, *Topology optimization: theory, methods, and applications*, Springer Science & Business Media, 2013.

[6] A. Cherkaev and E. Cherkaeva, *Optimal design for uncertain loading condition*, in Homogenization: In Memory of Serguei Kozlov, World Scientific, 1999, pp. 193–213.

[7] M. Cuturi, *Sinkhorn distances: Lightspeed computation of optimal transport*, Advances in neural information processing systems, 26 (2013).

[8] C. Dapogny and F. Feppon, *Shape optimization using a level set based mesh evolution method: an overview and tutorial*, in preparation, (2022).

[9] 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.

[10] R. Gao and A. J. Kleywegt, *Distributionally robust stochastic optimization with wasserstein distance*, arXiv preprint arXiv:1604.02199, (2016).

[11] F. Hecht, *New development in freefem++*, Journal of numerical mathematics, 20 (2012), pp. 251–266.

[12] F. Lin, X. Fang, and Z. Gao, *Distributionally robust optimization: A review on theory and applications*, Numerical Algebra, Control & Optimization, 12 (2022), p. 159.

[13] K. Maute, *Topology optimization under uncertainty*, in Topology optimization in structural and continuum mechanics, Springer, 2014, pp. 457–471.

[14] Q. Merigot and B. Thibert, *Optimal transport: discretization and algorithms*, in Handbook of Numerical Analysis, vol. 22, Elsevier, 2021, pp. 133–212.

[15] P. Mohajerin 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.

[16] 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.

[17] H. Rahimian and S. Mehrotra, *Distributionally robust optimization: A review*, arXiv preprint arXiv:1908.05659, (2019).

[18] F. Santambrogio, *Optimal transport for applied mathematicians*, Birkäuser, 2015.

[19] A. Shapiro, D. Dentcheva, and A. Ruszczynski, *Lectures on stochastic programming: modeling and theory*, SIAM, 2021.

[20] J. Wang, R. Gao, and Y. Xie, *Sinkhorn distributionally robust optimization*, arXiv preprint arXiv:2109.11926, (2021).

[21] J. Zhen, D. Kuhn, and W. Wiesemann, *Mathematical foundations of robust and distributionally robust optimization*, arXiv preprint arXiv:2105.00760, (2021).