Entropy regularized Wasserstein distance based distributionally robust shape and topology optimization

Charles Dapogny[1], Franck Iutzeler[1], Julien Prando[1] and Boris Thibert[1]

[1] Laboratoire Jean Kuntzmann, Université Joseph Fourier, Grenoble, France

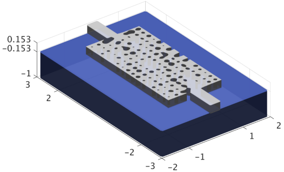
5th June, 2023

Foreword: optimal design and robustness

- The mathematical description of realistic systems involves physical parameters, e.g.
  - *In structure mechanics:* loads, elastic coefficients.
  - *In fluid mechanics:* viscosity, density of the fluid.
- These are often known imperfectly, through measurements, because either
  - They are measured or estimated,
  - They are altered during the use of the design.
- The optimality of a design is very sensitive to the parameters describing its environment,

  $\Rightarrow$ Need for "Robust" optimal design.
- All the formulations of this requirement suffer from drawbacks.
- The idea of distributional robustness is a remedy to the main conceptual flaw of stochastic approaches.

Turbine blades operate under very uncertain load and temperature conditions.

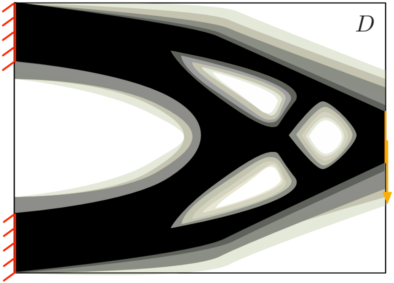


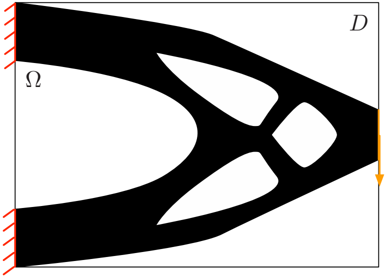
The wavelength of light injected into nanophotonic components is uncertain.

Contents

A generic, abstract optimal design setting (I)

- The design h is sought within a set $\mathcal{U}_{\text{ad}}$:
  - $h : D \to [0, 1]$ may be a "grayscale" density function, defined on a large "hold-all" domain D;
  - h may be a "black-and-white" shape $\Omega \subset \mathbb{R}^d$.
- The physical parameters are aggregated into an element ξ in a set $\Xi \subset \mathbb{R}^k$,
  When h is an elastic structure, ξ may represent the loads applied on h, or the material parameters (Young's modulus, Poisson's ratio).
- The cost of h when the parameters ξ are at play is $\mathcal{C}(h, \xi)$.
  When h is a structure, $\mathcal{C}(h, \xi)$ may be the compliance of h, depending on the elastic displacement $u_{h,\xi}$.





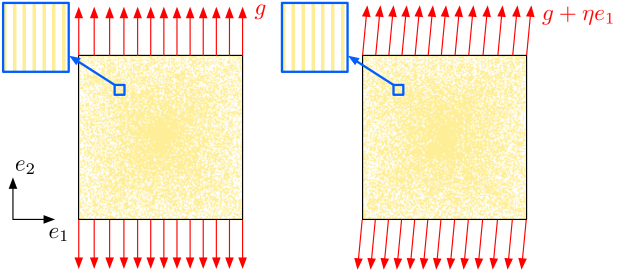
A generic, abstract optimal design setting (II)

- The nominal optimal design problem, when ξ is known exactly, reads:

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

  where constraints are omitted for simplicity.

- The optimal character of a design is strongly dependent on the actual value of ξ.



The optimal elastic microstructure to cope with a vertical traction load ξ for the compliance $\mathcal{C}(h, \xi)$ is also the worst one when an arbitrarily small horizontal component is added to ξ: $\mathcal{C}(h, \xi + \eta \mathbf{e_1}) = \infty$ [CheChe].

$\Rightarrow$ Need for a means to incorporate "robustness" into the optimal design problem.

The worst-case design approach

When no information is available about ξ except for a maximum bound m, robustness with respect to ξ is usually enforced via the worst-case program:

$$\min_{h \in \mathcal{U}_{\mathrm{ad}}} J_{\mathrm{wc}}(h), \text{ where } J_{\mathrm{wc}}(h) = \sup_{||\xi|| \leq m} \mathcal{C}(h, \xi).$$

Drawbacks of this approach:

- This yields a difficult and costly bi-level optimization problem, which can often be addressed only via coarse approximations
- Such formulations often lead to "pessimistic" designs, showing poor nominal performance for the sake of anticipating an unlikely worst-case scenario.

Stochastic approaches (I)

- Stochastic approaches assume the knowledge of the law of ξ, as a probability measure $\mathbb{P}_{\text{true}} \in \mathcal{P}(\Xi)$:

$$\forall A \subset \Xi, \text{ the probability that } \xi \text{ belong to } A \text{ is } \mathbb{P}_{\text{true}}(\{\xi \in A\}) = \int_A \mathrm{d}\mathbb{P}_{\text{true}}(\xi).$$

- Stochastic optimal design problems involve a probabilistic quantity of $\mathcal{C}(h, \xi)$, e.g.
  - The mean value

$$E(h) := \int_\Xi \mathcal{C}(h, \xi)\, \mathrm{d}\mathbb{P}_{\text{true}}(\xi).$$

  - Other quantiles such as the variance

$$V(h) = \int_\Xi \left(\mathcal{C}(h, \xi) - E(h)\right)^2 \mathrm{d}\mathbb{P}_{\text{true}}(\xi).$$

  - A probability of failure:

$$P(h) := \mathbb{P}_{\text{true}}\left(\{\xi \in \Xi \text{ s.t. } \mathcal{C}(h, \xi) > C_T\}\right),$$

  where C_T is a safety threshold.

Stochastic approaches (II)

Drawbacks of this approach:

- The evaluation of the stochastic integrals at play is very costly.
- The probability law $\mathbb{P}_{\text{true}}$ of ξ is assumed to be known, while often, this law is only accessible through a set of samples ξ_i, $i = 1, \dots, N$.

The recent idea of distributionally robustness alleviates the need for an exact knowledge of the law of ξ, and demands only an "estimate" $\mathbb{P}$ of the latter:

$$\text{Minimize the worst value } \sup_{\mathbb{Q} \text{ "close" to } \mathbb{P}} \int_{\Xi} \mathcal{C}(h, \xi) \, \mathrm{d}\mathbb{Q}(\xi) \text{ of the mean value of } \mathcal{C}(h, \xi)$$

when the uncertainty law $\mathbb{Q} \in \mathcal{P}(\Xi)$ is "close" to $\mathbb{P}$.

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.

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

The distributionally robust approach (I)

- The only available information about the uncertainty $\xi \in \Xi$ is a nominal law $\mathbb{P}$, which is e.g. the empirical mean of observed samples ξ_i, $i = 1, \dots, N$:

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

- The distributionally robust optimal design problem reads:

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

where the ambiguity set $\mathcal{U}(\mathbb{P}) \subset \mathcal{P}(\Xi)$ is made of the probability laws $\mathbb{Q}$ which are "close" to $\mathbb{P}$.

The distributionally robust approach (II)

- The ambiguity set $\mathcal{U}(\mathbb{P})$ may be of various natures:
  - $\mathcal{U}(\mathbb{P})$ may be the set of probability measures on Ξ whose moments are "close" to those of $\mathbb{P}$, up to a certain order k:

$$\mathcal{U}(\mathbb{P}) = \left\{ \mathbb{Q} \in \mathcal{P}(\Xi) \text{ s.t. } \sup_{|\alpha| \leq k} \left| \int_\Xi \xi^\alpha \, \mathrm{d}\mathbb{P}(\xi) - \int_\Xi \xi^\alpha \, \mathrm{d}\mathbb{Q}(\xi) \right| \leq m \right\}.$$

  - $\mathcal{U}(\mathbb{P})$ may be the set of probability measures which are "close" to $\mathbb{P}$,

$$\mathcal{U}(\mathbb{P}) = \left\{ \mathbb{Q} \in \mathcal{P}(\Xi), \; d(\mathbb{P}, \mathbb{Q}) \leq m \right\},$$

    where $d(\cdot,\cdot)$ is a suitable notion of "distance" on $\mathcal{P}(\Xi)$.

- Depending on $\mathcal{U}(\mathbb{P})$, the problem may be amenable to a tractable reformulation.

- We shall consider ambiguity sets of the latter type, relying on the entropy-regularized Wasserstein distance on $\mathcal{P}(\Xi)$.

The Wasserstein distance (I)

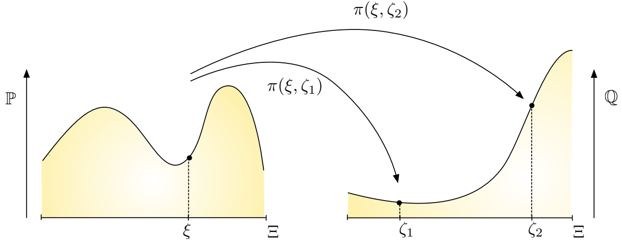
- A coupling is a probability measure $\pi \in \mathcal{P}(\Xi \times \Xi)$.
- The first and second marginals $\pi_1, \pi_2 \in \mathcal{P}(\Xi)$ of $\pi \in \mathcal{P}(\Xi \times \Xi)$ are defined by:

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

$$\int_{\Xi\times\Xi} \varphi(\zeta) \mathrm{d}\pi(\xi,\zeta) = \int_{\Xi} \varphi(\zeta)\, \mathrm{d}\pi_2(\zeta).$$

- *Interpretation:* If $\pi \in \mathcal{P}(\Xi \times \Xi)$ is a coupling with marginals $\mathbb{P}, \mathbb{Q} \in \mathcal{P}(\Xi)$,

$\pi(\xi,\zeta) \approx$ proportion of the mass of $\mathbb{Q}$ at ζ coming from the mass at ξ in $\mathbb{P}$.



$\pi(\xi,\zeta)$ is the amount of mass at ζ in the second marginal $\mathbb{Q}$ coming from the mass at ξ in the first marginal $\mathbb{P}$.

The Wasserstein distance

Definition 1.

Let Ξ be a compact subset of $\mathbb{R}^k$; the Wasserstein distance $W(\mathbb{P},\mathbb{Q})$ between two probability measures $\mathbb{P}$, $\mathbb{Q} \in \mathcal{P}(\Xi)$ is

$$W(\mathbb{P},\mathbb{Q}) = \inf_{\substack{\pi\in\mathcal{P}(\Xi\times\Xi)\\ \pi_1=\mathbb{P},\ \pi_2=\mathbb{Q}}} \int_{\Xi\times\Xi} c(\xi,\zeta)\,\mathrm{d}\pi(\xi,\zeta),$$

where the ground cost $c(\xi,\zeta)$ on $\mathbb{R}^d$ is chosen to be quadratic $c(\xi,\zeta) = |\xi-\zeta|^2$.

- The Wasserstein distance "lifts" the ground cost $c(\xi,\zeta)$ on Ξ into a distance over measures on Ξ.
- It is a flexible means to evaluate the distance between $\mathbb{P},\mathbb{Q} \in \mathcal{P}(\Xi)$, which can smoothly identify differences (e.g. translations) on the supports of $\mathbb{P}$ and $\mathbb{Q}$.

G. Peyré and M. Cuturi, *Computational optimal transport: With applications to data science*, Foundations and Trends in Machine Learning, 11 (2019), pp. 355–607.

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

Entropy-regularization of the Wasserstein distance

- For a variety of purposes, this quantity is often regularized [Cu]:

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

where the entropy $H(\pi)$ of a coupling π is:

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

- The fixed reference coupling $\pi_0 \in \mathcal{P}(\Xi\times\Xi)$ plays the role of a "prior".
- A judicious choice about π_0, with nice statistical guarantees, is

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

for some $\sigma > 0$ and a normalization factor α_ξ [AlMa]:

$$\text{For all } \varphi \in \mathcal{C}(\Xi\times\Xi), \quad \int_{\Xi\times\Xi} \varphi(\xi,\zeta)\,\mathrm{d}\pi_0(\xi,\zeta) = \int_\Xi \left(\int_\Xi \varphi(\xi,\zeta)\,\mathrm{d}\nu_\xi(\zeta) \right) \mathrm{d}\mathbb{P}(\xi).$$

- *Intuition:* π_0 "spreads" the mass of $\mathbb{P}$ at ξ over a characteristic length scale σ.

The distributionally robust optimal design problem

The entropy-regularized distributionally robust optimal design problem reads:

$$\min_{h \in \mathcal{U}_{\mathrm{ad}}} J_{\mathsf{dr},\varepsilon}(h), \text{ where } J_{\mathsf{dr},\varepsilon}(h) = \sup_{\substack{\mathbb{Q} \in \mathcal{P}(\Xi) \\ W_\varepsilon(\mathbb{P},\mathbb{Q}) \leq m}} \int_\Xi \mathcal{C}(h,\xi)\, \mathrm{d}\mathbb{Q}(\xi).$$

This problem looks very difficult to treat at first glance... but it can be given a tractable reformulation up to the use of convex duality.

A convex duality result (I)

Proposition 1.

Besides "mild" assumptions, suppose that:

- Ξ *is a convex and compact subset of* $\mathbb{R}^k$,
- $f : \Xi \to \mathbb{R}$ *is a continuous function,*
- $\mathbb{P} \in \mathcal{P}(\Xi)$ *is a probability measure.*

For any $m > 0$, *and for a sufficiently small value of* σ, *the following equality holds:*

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

Intuition: We introduce a Lagrange multiplier λ for the constraint on $W_\varepsilon(\mathbb{P},\mathbb{Q})$:

$$\begin{aligned} \sup_{W_\varepsilon(\mathbb{P},\mathbb{Q}) \le m} \int_\Xi f(\zeta)\, \mathrm{d}\mathbb{Q}(\zeta) &= \sup_{\mathbb{Q} \in \mathcal{P}(\Xi)} \inf_{\lambda \ge 0} \left(\int_\Xi f(\zeta)\, \mathrm{d}\mathbb{Q}(\zeta) + \lambda(m - W_\varepsilon(\mathbb{P},\mathbb{Q})) \right), \\ &= \inf_{\lambda \ge 0} \sup_{\mathbb{Q} \in \mathcal{P}(\Xi)} \left(\int_\Xi f(\zeta)\, \mathrm{d}\mathbb{Q}(\zeta) + \lambda(m - W_\varepsilon(\mathbb{P},\mathbb{Q})) \right), \end{aligned}$$

where the exchange of the infimum and supremum proceeds from convex duality.

A convex duality result (II)

Inserting the definition of $W_\varepsilon(\mathbb{P},\mathbb{Q})$, it follows:

$$\sup_{W_\varepsilon(\mathbb{P},\mathbb{Q})\le m} \int_\Xi f(\zeta)\mathrm{d}\mathbb{Q}(\zeta) = \inf_{\lambda\ge 0} \sup_{\substack{\pi\in\mathcal{P}(\Xi\times\Xi)\\ \pi_{\mathbf{1}}=\mathbb{P}}} \left\{\lambda m + \int_\Xi \Big(f(\zeta) - \lambda c(\xi,\zeta) - \lambda\varepsilon H(\pi)\Big)\,\mathrm{d}\pi(\xi,\zeta)\right\},$$

Given the definition of $H(\pi)$, the maximization holds over couplings π of the form

$$\pi(\xi,\zeta) = a(\xi,\zeta)\pi_0(\xi,\zeta),\ \text{ for some function } a \in L^1(\Xi\times\Xi),$$

and so:

$$\sup_{W_\varepsilon(\mathbb{P},\mathbb{Q})\le m} \int_\Xi f(\zeta)\,\mathrm{d}\mathbb{Q}(\zeta) = \inf_{\lambda\ge 0} \sup_{\substack{\alpha\in L^{\mathbf{1}}(\Xi\times\Xi)\\ \int_\Xi \alpha(\xi,\zeta)\,\mathrm{d}\nu_\xi(\zeta)=\mathbf{1}}} \Bigg\{\lambda m$$

$$+\int_\Xi \Big(f(\zeta) - \lambda c(\xi,\zeta) - \lambda\varepsilon \log\alpha(\xi,\zeta)\Big)\alpha(\xi,\zeta)\,\mathrm{d}\pi_0(\xi,\zeta)\Bigg\}.$$

Exploiting the Euler-Lagrange equation for the inner maximization, we obtain:

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

and the desired result follows.

A convex duality result (III)

- The entropy-regularized distributionally robust optimization problem rewrites:

$$\min_{\substack{h \in \mathcal{U}_{\text{ad}} \\ \lambda \geq \mathbf{0}}} \mathcal{D}(h, \lambda), \text{ where}$$

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

- This problem can be solved by a standard optimization algorithm based on the derivatives of the objective functional $\mathcal{D}(h, \lambda)$ with respect to h and λ.

- Constraints could be added to the problem without additional conceptual difficulty.

Topology optimization of a 2d mast (I)

- We consider the topology optimization of a 2d mast.
- The uncertain parameter ξ is the (constant) load vector applied on the two arms Γ_N.
- The considered problem is

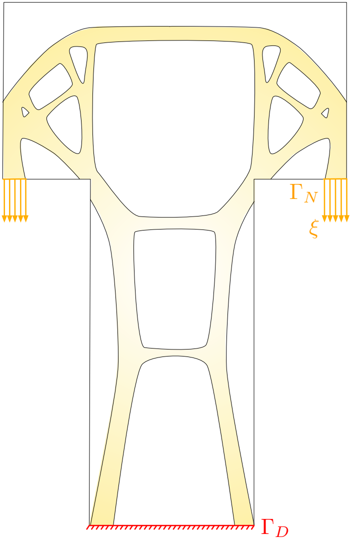
$$\min_{h \in \mathcal{U}_{\mathrm{ad}}} J_{\mathrm{dr},\varepsilon}(h) \text{ s.t. } \mathrm{Vol}(h) = V_T,$$

built from the compliance as the cost $\mathcal{C}(h,\xi)$:

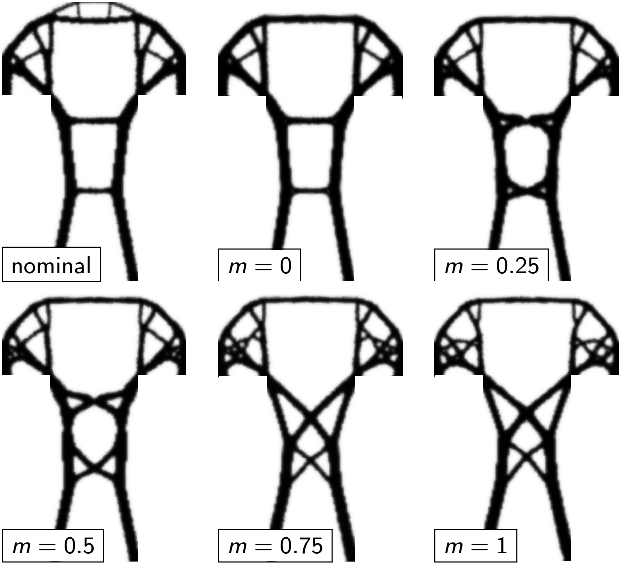
$$\mathcal{C}(h,\xi) = \int_D \xi \cdot u_{h,\xi} \, \mathrm{d}x.$$

- The nominal law is made from a single observation

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



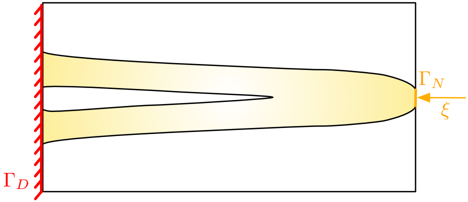
Topology optimization of a 2d mast (II)



Optimized density in the mast topology optimization example.

Shape optimization of a 2d cantilever (I)

- We consider the shape optimization of a 2d cantilever.
- The uncertain parameter ξ is the (constant) load vector applied on Γ_N.



- The considered problem is

$$\min_{\Omega} J_{\mathrm{dr},\varepsilon}(\Omega) \text{ s.t. } \mathrm{Vol}(\Omega) = V_T,$$

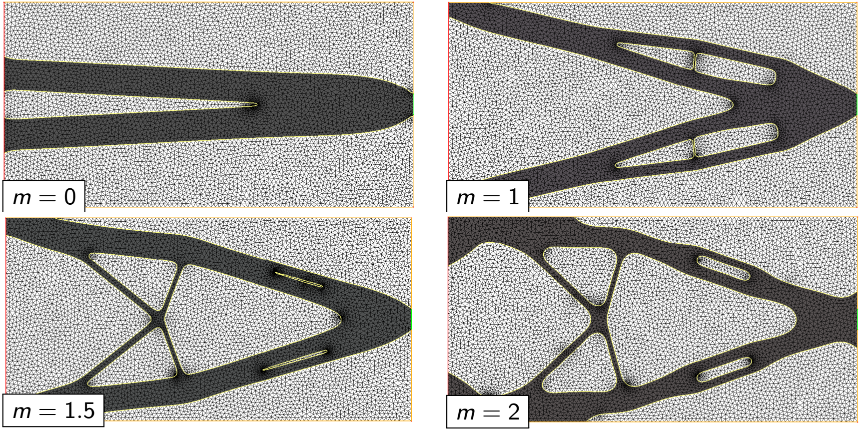
where the cost $\mathcal{C}(\Omega, \xi)$ is the compliance:

$$\mathcal{C}(\Omega, \xi) = \int_{\Omega} \xi \cdot u_{\Omega,\xi} \, \mathrm{d}x.$$

- The nominal law is made from a single observation

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

Shape optimization of a 2d cantilever (II)



Optimized shape in the 2d cantilever shape optimization example.

Conclusion and perspectives

Conclusions:

- Distributional robustness considers the minimization of the worst case

$$h \mapsto \sup_{\mathbb{Q} \text{ ``close'' to } \mathbb{P}} \mathbb{E}_{\xi \sim \mathbb{Q}}(\mathcal{C}(h, \xi))$$

  of the mean value of a cost $\mathcal{C}(h, \xi)$ when the law $\mathbb{Q}$ of ξ is close to a nominal law $\mathbb{P}$.

- One relevant notion of distance between probability laws is the (entropy-regularized) Wasserstein distance.

- The distributionally robust optimization problem can be given a tractable reformulation owing to convex duality.

Perspectives:

- Application of this methodology to more realistic problems, e.g. with nominal laws built from multiple samples.

- Distributionally robust versions of other probabilistic quantities of the cost: variance, quantiles, conditional value at risk, ...

Thank you !

Thank you for your attention!

References I

[AdDaFre] G. Allaire, C. Dapogny, and P. Frey, *Shape optimization with a level set based mesh evolution method*, Comput Methods Appl Mech Eng, 282 (2014), pp. 22–53.

[AIMa] W. Azizian, F. Iutzeler and J. Malick, *Regularization for wasserstein distributionally robust optimization*, ESAIM: Control, Optimization and Calculus of Variations, 29, (2023),33.

[CheChe] A. Cherkaev and E. Cherkaeva, *Optimal design for uncertain loading condition*, Homogenization: In Memory of Serguei Kozlov. World Scientific, (199), pp. 193–213.

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

[DaIMeThi] C. Dapogny, F. Iutzeler, A. Meda and B. Thibert, *Entropy-regularized Wasserstein distributionally robust shape and topology optimization*, Struct. Multisdic. Optim., 66(3), (2023), p. 42.

References II

[EsKu] 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.

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

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

[PeyCu] G. Peyré and M. Cuturi, *Computational optimal transport: With applications to data science*, Foundations and Trends in Machine Learning, 11 (2019), pp. 355–607.

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