First-Order Methods
for Distributionally Robust Mixed-Integer Optimization

Hubert VILLUENDAS[1]

Université Grenoble Alpes, France,
hubert.villuendas@univ-grenoble-alpes.fr

Keywords : *Data driven optimization, robust optimization, data uncertainty.*

Context. We consider a general class of combinatorial optimization problems in which some imput data are only partly observed or subject to estimation errors. The problem can be written as a function of a scenario $\xi \in \Xi$:

$$\underset{z \in \mathcal{Z}}{\text{minimize}} \quad f(z, \xi),$$

where $\mathcal{Z}$ is the mixed-integer set of decision variables and $f(z,\xi)$ is the optimal value of the combinatorial problem under scenario ξ for $\Xi \subseteq \mathbb{R}^d$ the set of all possible scenarios of the uncertain data. For example, in a facility location problem with uncertain demand (FIG. 1), $\mathcal{Z}$ would be the set of facility opening vectors and $f(z,\xi)$ the optimal value for the opening z under demand ξ, which can itself be a convex program.

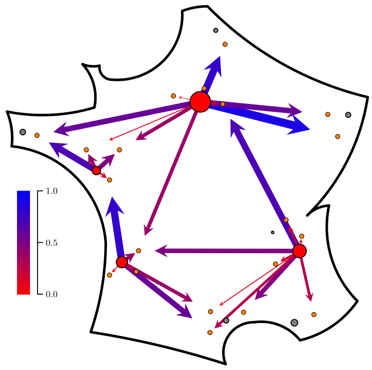

FIG. 1: Facility location under demand uncertainty: openings z ●, and allocations (customers arrows).

Assuming that the data follows an unknown probability distribution $\mathbb{P}$, the ideal goal is to choose z to minimize the expected value $\mathbb{E}_{\xi\sim\mathbb{P}}[f(z,\xi)]$. When the probability distribution of uncertain parameters is unknown, one typically relies on a finite historical dataset $\{\widehat{\xi}_1, \ldots, \widehat{\xi}_N\} \subset \Xi$ to infer plausible distributions. Using the empirical measure $\widehat{\mathbb{P}}_N \overset{\text{def}}{=} \frac{1}{N}\sum_{i=1}^N \delta_{\widehat{\xi}_i}$, data driven models typically relies on empirical risk minimization

$$\min_{z\in\mathcal{Z}} \frac{1}{N}\sum_{i=1}^{N} f\left(z, \widehat{\xi}_i\right). \tag{ERM}$$

However, this approach can fail in the case where N is small or when the deployment data differs from the training data. In this context, Wasserstein Distributionally Robust Optimization (WDRO) has become a powerful framework, by its nice modeling and generalisation properties [1, 2, 3]: we minimize the expected value of $f(z,\cdot)$ under the worst-case distribution within a neighborhood of the empirical measure $\widehat{\mathbb{P}}_N$. This leads to the problem:

$$\min_{z\,\in\,\mathcal{Z}} \sup_{\substack{\mathbb{Q}\,\in\,\text{Proba}(\Xi)\\ W_c(\mathbb{Q},\widehat{\mathbb{P}}_N)\,\leq\,\varrho}} \mathbb{E}_{\zeta\sim\mathbb{Q}}\left[f(z,\zeta)\right] \tag{WDRO}$$

where $W_c(\cdot,\cdot)$ denotes the Wasserstein distance [4], defined as

$$W_c\left(\mathbb{Q}_1, \mathbb{Q}_2\right) \overset{\text{def}}{=} \inf_{\substack{\pi\,\in\,\text{Proba}(\Xi\times\Xi)\\ [\pi]_1=\mathbb{Q}_1\;[\pi]_2=\mathbb{Q}_2}} \int_{\Xi\times\Xi} c\left(\xi,\xi'\right)\mathrm{d}\pi\left(\xi,\xi'\right).$$

Approach. Instead of (WDRO), we propose to use its entropic-regularized version, following [5]:

$$\min_{z\in\mathcal{Z}}\min_{\lambda\geq 0}\left[F(z,\lambda)\stackrel{\text{def}}{=}\lambda\varrho+\varepsilon\mathbb{E}_{\xi\sim\widehat{\mathbb{P}}_N}\left[\log\left(\mathbb{E}_{\zeta\sim\mathcal{N}(\xi,\sigma^2\mathbf{I})}\left[\exp\left(\frac{f(z,\zeta)-\lambda c(\xi,\zeta)}{\varepsilon}\right)\right]\right)\right]\right].$$

where $\lambda\geq 0$ is the dual variable associated to the Wasserstein distance constraint. This problem retains the generalization and robustness guarantees of the original formulation and enables the computation of stochastic gradient estimators via a Monte-Carlo sampling method (see FIG. 2) [6].

Then, we use a stochastic Frank-Wolfe algorithm to handle sampling noise [7] which preserves the combinatorial nature of the feasible set while ensuring scalability. Our approach is close to oracle-based methods for robust optimization, see *e.g.* [8], and is complementarty to existing WDRO approaches for mixed-integer problems that rely on exact reformulations combined with branch-and-bound or cutting-plane methods [3]. Our use of a stochastic Frank–Wolfe algorithm enables an oracle-based optimization over conv($\mathcal{Z}$), avoiding an explicit description of the feasible set and allowing efficient treatment of rich combinatorial structures.

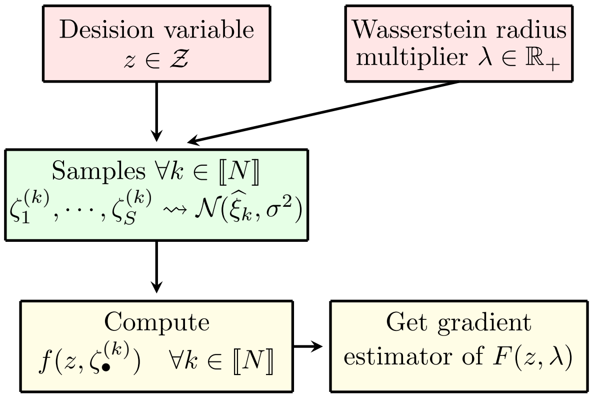

FIG. 2: Gradient computation scheme.

This framework thus extends the applicability of WDRO methods to mixed-integer and discrete optimization problems, offering a principled and computationally efficient approach to decision-making under uncertain data.

References

[1] Daniel Kuhn and Peyman Mohajerin Esfahani. *Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations.* Mathematical Programming, (2018).

[2] Tam Le and Jérôme Malick. *Universal generalization guarantees for Wasserstein distributionally robust models.* ICLR, (2025).

[3] Wolfram Wiesemann and Daniel Kuhn. *Distributionally robust convex optimization.* Operations research, INFORMS, (2014).

[4] Gabriel Peyré and Marco Cuturi. *Computational Optimal Transport.* Foundations and Trends in Machine Learning, (2019).

[5] Waïss Azizian, Franck Iutzeler and Jérôme Malick. *Regularization for wasserstein distributionally robust optimization.* ESAIM: Control, Optimisation and Calculus of Variations, (2023).

[6] Florian Vincent, Waïss Azizian, Franck Iutzeler and Jérôme Malick. `skwdro`*: a library for Wasserstein distributionally robust machine learning.* arXiv preprint, (2024).

[7] Gábor Braun et al. *Conditional Gradient Methods* (2025).

[8] Mathieu Besançon and Jannis Kurtz. *A Frank-Wolfe Algorithm for Oracle-based Robust Optimization.* arXiv preprint, (2024).