Discrete Optimization: Theory, applications and computation

Agostinho Agra
CEMS.UL and DCM, University of Lisbon

Universidade de Lisboa

January 27, 2026

Outline

Background

2004 - PhD (Mixed Integer Programming)

2006 - CORE 40th anniversary

2010 –> Robust Optimization and other approaches to deal with uncertainty (stochastic, CVaR)

2020 –> Distributionally Robust Optimization

Modeling uncertainty

- Stochastic Program

$$\min_{x \in X} \{ c^t x + E_Q(R(x, \xi)) \}$$

- Robust Optimization

$$\min_{x \in X} \left\{ c^t x + \max_{\omega \in \Omega} R(x, \omega) \right\}$$

- Distributionally Robust Optimization

$$\min_{x \in X} \left\{ c^t x + \max_{P \in \mathbb{A}} E_P(R(x, \xi)) \right\}$$

Distributionally Robust Optimization

Ambiguity set

The Distributionally Robust Optimization assumes that the underlying probability distribution is unknown and lies in an ambiguity set of probability distributions, known as *ambiguity set*

As in robust optimization, this approach hedges against the ambiguity in probability distribution by taking a worst-case approach.

- A common approach to define the ambiguity set is to assume that the true probability close to a reference / nominal distribution
- The reference distribution is obtained from the historical data
- A distance function must be defined

"The ambiguity set should also be easy to parameterize from data, and - ideally - it should facilitate a tractable reformulation of the distributionally robust optimization problem as a structured mathematical program that can be solved with off-the-shelf optimization software."

Esfahani and Kuhn. Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations. Mathematical Programming, 171(1):115–171, 2018.

Ambiguity set

Some tractable cases assume the random vector ξ has finitely many realizations:

- Moment matching set: bounds are imposed to the moments
- Kantorovich or 1-norm Wasserstein ambiguity set

M. Bansal, K.-L. Huang, and S. Mehrotra. Decomposition algorithms for two-stage distributionally robust mixed binary programs. SIAM Journal on Optimization, 28(3):2360–2383, 2018.

Kantorovich or 1-norm Wasserstein ambiguity set

Support $\Omega = \{1, \ldots, |\,\Omega\,|\}$

Q_ω represents the reference (nominal) distribution - obtained from the data

The decision variables k give the transportation plan for moving a mass distribution described by Q to another one described by P.

$k_{\omega\omega'}$ is the amount to transport from Q_ω to $P_{\omega'}$

The Wasserstein distance between Q and P represents the cost of an optimal mass transportation plan.

$$\sum_{\omega \in \Omega} \sum_{\omega' \in \Omega} \|\, \omega - \omega' \,\|_1 \; k_{\omega\omega'}$$

Kantorovich or 1-norm Wasserstein ambiguity set

$$
\begin{aligned}
\mathbb{A}(\epsilon) = \Big\{ P \in \mathbb{R}^{|\Omega|} : & \sum_{\omega \in \Omega} \sum_{\omega' \in \Omega} \| \omega - \omega' \|_1 \, k_{\omega\omega'} \leq \epsilon \\
& \sum_{\omega' \in \Omega} k_{\omega\omega'} = P_\omega, && \omega \in \Omega \\
& \sum_{\omega \in \Omega} k_{\omega\omega'} = Q_{\omega'}, && \omega' \in \Omega \\
& \sum_{\omega \in \Omega} P_\omega = 1 \\
& P_\omega \geq 0, && \omega \in \Omega \\
& k_{\omega\omega'} \geq 0, && \omega, \omega' \in \Omega \Big\}
\end{aligned}
$$

Kantorovich ambiguity set

- When $\epsilon = 0$ then variables $k_{\omega\omega'}$ are null and we obtain a singleton $\{Q\} \Rightarrow$ Stochastic Program

$$\min_{x \in X} \left\{ c^t x + \max_{P \in \mathbb{A}} \sum_{\omega \in \Omega} P_\omega R(x, \omega) \right\} \rightarrow \min_{x \in X} \left\{ c^t x + \sum_{\omega \in \Omega} Q_\omega R(x, \omega) \right\}$$

- When ϵ is large enough all the mass probability can be concentrated in one scenario $\Rightarrow$ Robust Optimization

$$\min_{x \in X} \left\{ c^t x + \max_{P \in \mathbb{A}} \sum_{\omega \in \Omega} P_\omega R(x, \omega) \right\} \rightarrow \min_{x \in X} \left\{ c^t x + \max_{\omega \in \Omega} R(x, \omega) \right\}$$

Pareto front for two-stage distributionally robust optimization problems
A Agra, F Rodrigues. European Journal of Operational Research, 2025

How to interpret and compute ϵ?

- Esfahani and Kuhn (2018) - assume the existence of a large sample from the (unknown) true probability distribution, and such a sample is then used to build training samples and conduct out-of-sample analysis. The value of ϵ can then be estimated from the out-of-sample analysis
- Zhao and Guan (2018) - ϵ is computed from the sample size and for a given confidence level for the ambiguity set

How to interpret and compute ϵ?

"Our" view

ϵ may be used to control the degree of uncertainty the decision-maker wants to consider

A small value of ϵ means the decision-maker wishes to consider scenarios closer to the historical data, while large values of ϵ lead to scenarios where only the support is known.

Idea

Obtain tradeoffs between degrees of uncertainty and costs/profits - similar to Pareto front

$$\min_{x \in X} \left\{ c^\top x + \max_{p \in \mathbb{A}_\epsilon} \mathbb{E}_p[R(x, \xi)] \right\}$$

The function $R(x, \xi)$ is known as the recourse function, which gives the second-stage cost for each x.

Given an interval $[0, L]$, the continuous trajectory of the first-stage solution $x \in X$ is the function

$$z_x(\epsilon) = c^\top x + \max_{p \in \mathbb{A}_\epsilon} \mathbb{E}_p[R(x, \xi)]$$

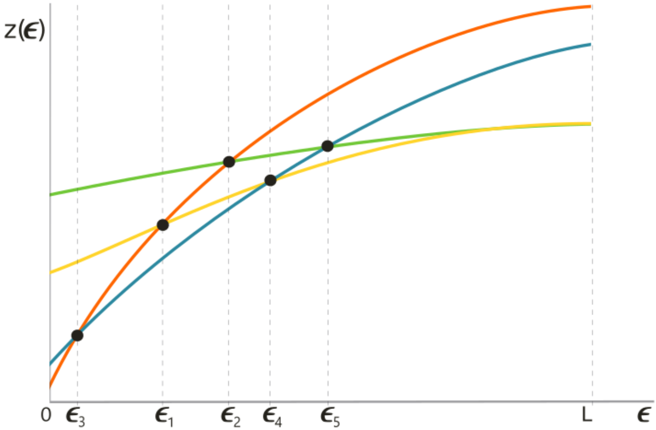


Figure: Partial illustration of several trajectories

Solving DRO

How to solve DRO problems with finite support?

Two-stage distributionally robust optimization with a finite support
A Agra, Computers & Operations Research, 107142, 2025

Distributionally robust model

$$\min_{x \in X} \left\{ c^t x + \max_{p \in \mathbb{A}} \sum_{\omega \in \Omega} P_\omega R(x, \omega) \right\}$$

where

- x first stage solution
- $\mathbb{A}$ Ambiguity set
- Ω Support of the probability distributions
- $R(x, \omega)$ Recourse function

Assume:

$$R(x, \omega_i) = \min\{g_i^\top y_i : W_i y_i \geq d_i - T_i x\}$$

$$\Omega = N = \{1, \ldots, n\}$$

Ambiguity set

$$
\begin{aligned}
\mathbb{A} = \Bigg\{ p \in \mathbb{R}^n : \sum_{i\in N}\sum_{j\in N} a_{ij} t_{ij} \leq \epsilon \\
\sum_{i\in N} t_{ij} = p_j, \qquad & j \in N \\
\sum_{j\in N} t_{ij} = q_i, \qquad & i \in N \\
p_j \geq 0, \qquad & j \in N \\
t_{ij} \geq 0, \qquad & i,j \in N \Big\}
\end{aligned}
$$

where $a_{ij} \geq 0$

$a_{ij} = \| \omega_i - \omega_j \|_1$ - Kantorovich ambiguity set

Solution procedure: L-shape method

$$\min\ c^\top x + \theta \tag{1}$$

$$\text{s.t.}\quad \theta \geq \max_{p \in \mathbb{A}} \sum_{i \in N} p_i R(x, \omega_i) \tag{2}$$

$$x \in X \tag{3}$$

Follow Benders decomposition:

$$R(x, \omega_i) = \min\{g_i^\top y_i : W_i y_i \geq d_i - T_i x\}$$

$$= \max\{\pi_i(x)^\top (d_i - T_i x) | \pi_i(x)^\top W_i \leq g_i, \pi_i(x) \geq 0\}$$

Let:
x' - first-solution x',
$\pi_i(x')$ - optimal dual multipliers for $R(x', \omega_i)$
$p_i^*(x')$ - optimal solution for the transportation problem with objective function coefficients $R(x', \omega_i), i \in N$

Cut:

$$\theta \geq \sum_{i \in N} p_i^*(x')(\pi_i(x')^\top (d_i - T_i x)). \tag{4}$$

Master Problem:

$$\begin{aligned} \min\ & c^\top x + \theta \\ \text{s.t.}\ & \theta \geq \sum_{i \in N} p_i^*(x^t)(\pi_i(x^t)^\top (d_i - T_i x)), \forall t \in \{1, \dots, r\} \\ & x \in X \end{aligned}$$

Solve Master Problem

Solve the second-stage subproblem $R(x, \omega)$ for each scenario ω

Solve the transportation problem to determine the probabilities $P\omega$

Add Cut $\theta \geq$ ....

Solution procedure: row-and-column generation

The second approach is also based on the epigraph formulation however, it differentiates from the L-shape approach in the way constraints $\theta \geq \max_{p \in \mathbb{A}} \sum_{i \in N} p_i R(x, \omega_i)$, are handled

$$\theta \geq \sum_{i \in N} p_i R(x, \omega_i), \forall p \in \mathbb{A}$$

As $R(x, \omega_i)$ is a minimization problem it suffices to replace $R(x, \omega_i)$ by the objective function of the recourse function and add the corresponding constraints:

$$\theta \geq \sum_{i \in N} p_i g_i^\top y_i, \forall p \in \mathbb{A}$$

$$W_i y_i \geq d_i - T_i x, i \in N$$

The full DRO model can be written as follows:

$$\min\ c^\top x + \theta \tag{5}$$

$$\text{s.t.}\ \ \theta \geq \sum_{i \in N} p_i g_i^\top y_i, \forall p \in \mathbb{A} \tag{6}$$

$$W_i y_i \geq d_i - T_i x, i \in N \tag{7}$$

$$x \in X \tag{8}$$

Set $\mathbb{A}$ includes an infinite number of distributions. Replace by $Ext(\mathbb{A})$, the extreme points that may be optimal to the subproblem

Generate cuts (6) and (7) dynamically

Solution procedure: dualization

Dualize the transportation problem

$$\max_{p \in \mathbb{A}} \sum_{i \in N} p_i r(x, \omega_i) = \min_{(\beta,\alpha) \in U} \sum_{i \in N} \beta_i q_i + \alpha\epsilon \tag{9}$$

where

$$U = \{(\beta, \alpha) | a_{ij}\alpha + \beta_i \geq r(x, \omega_j), \forall i, j \in N, \alpha \geq 0\}.$$

The DRO problem can be written as follows.

$$\begin{aligned}
\min\ & c^\top x + \sum_{i \in N} \beta_i q_i + \alpha\epsilon \\
\text{s.t.}\ \ & a_{ij}\alpha + \beta_i \geq g_j^\top y_j, \forall i, j \in N \\
& \alpha \geq 0 \\
& W_j y_j \geq d_j - T_j x, j \in N \\
& x \in X
\end{aligned}$$

Impact of ϵ on the scenarios

Solution procedures

Conclusions:

- The framework approach should be selected according to the ambiguity set characteristics
- The dualization approach is the simplest to apply and performs well except for the instances that are simultaneously harder and have large number of scenarios
- The row-and-column generation is suitable when the ambiguity set approximates the RO case
- The L-shape is the best approach for the remaining cases

Berth allocation

Joint work with Filipe Rodrigues (ISEG)

Goal

Mathematical models for the berth allocation problem

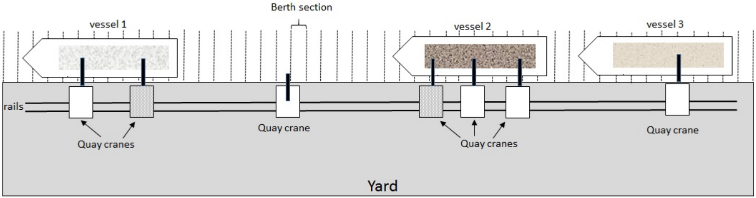
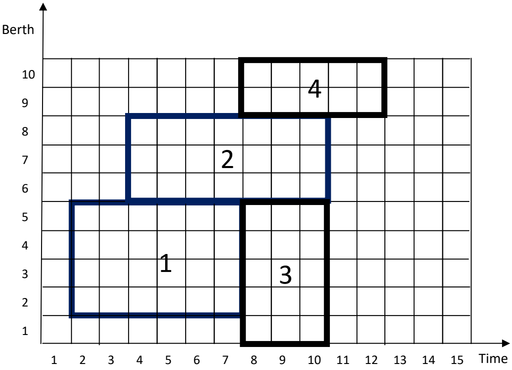


Figure: Example of a quay operation on three ships and seven cranes

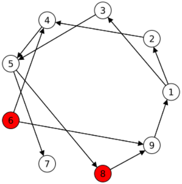

Berth
10
9
8
7
6
5
4
3
2
1
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Time
4
2
1
3

$$
\begin{aligned}
\min \quad & \sum_{k \in V} c_k && && (10)\\
s.t. \quad & x_{\ell k} + x_{k\ell} + y_{\ell k} + y_{k\ell} \geq 1, && k, \ell \in V, k < \ell, && (11)\\
& x_{\ell k} + x_{k\ell} \leq 1, && k, \ell \in V, k < \ell, && (12)\\
& y_{\ell k} + y_{k\ell} \leq 1, && k, \ell \in V, k < \ell, && (13)\\
& b_k \geq b_\ell + L_\ell + J(y_{k\ell} - 1), && k, \ell \in V, k \neq \ell, && (14)\\
& b_k \leq J - L_k, && k \in V, && (15)\\
& b_k \in \mathbb{Z}_0^+, && k \in V, && (16)\\
& x_{k\ell}, y_{k\ell} \in \{0, 1\}, && k, \ell \in V, k \neq \ell && (17)\\
& t_\ell \geq t_k + H_k + F + M(x_{k\ell} - 1), && k, \ell \in V, k \neq \ell, && (18)\\
& c_k \geq t_k + H_k - D_k, && k \in V, && (19)\\
& c_k \leq C, && k \in V, && (20)\\
& t_k \geq A_k, && k \in V, && (21)\\
& t_k, c_k \geq 0 && && (22)
\end{aligned}
$$

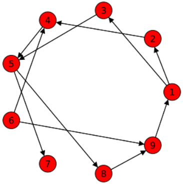
Interdiction models for graph propagation

Joint work with José Samuco (Univ. Aveiro)

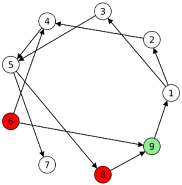
Given a graph $G = (V, E)$ and a set $S \subset V$ of activated/infected nodes, we aim to determining the set of c nodes that minimizes the network propagation on the subgraph that results from the removal of those c nodes.



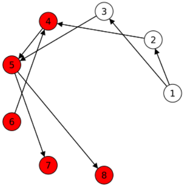
(a) The infected nodes are painted in red.



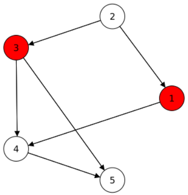
(b) Initial propagation. All nodes are infected.



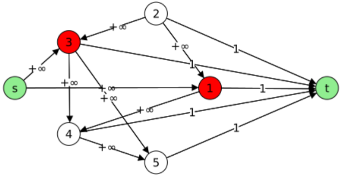
(c) The critical node is painted in green.



(d) Final propagation of infected nodes. Five nodes are infected.



(a) Original directed network with two infected nodes painted in red.



(b) The maximum flow network with source *s* and destination *t*.

Figure: Maximum flow network.

$$
\begin{aligned}
\min_{v} \quad & \max_{f} \sum_{i:(i,t)\in E} f(i,t) \\
\text{s.t.} \quad & \sum_{i\in V} v_i = c \\
& \sum_{i:(i,j)\in E} f(i,j) = \sum_{k:(j,k)\in E} f(j,k), \quad \forall j \in V \\
& f(i,j) \le \alpha(i,j)(1 - v_i), \quad \forall (i,j) \in E' \\
& f(i,j) \ge 0, \quad \forall (i,j) \in E' \\
& v_i \in \{0,1\}, \quad \forall i \in V.
\end{aligned}
$$

Thank you for your attention!

Acknowledgments

FCT within projects UID/04561/2025 DOI: https://doi.org/10.54499/UID/04561/2025