Robust Optimization : A tutorial

V. Leclère (ENPC)

May 21, 2019

Contents

Contents

An optimization problem

A generic optimization problem can be written

$$\begin{aligned} \min_x \quad & L(x) \\ s.t. \quad & g(x) \leq 0 \end{aligned}$$

where

- x is the decision variable
- L is the objective function
- g is the constraint function

An optimization problem with uncertainty

Adding uncertainty ξ in the mix

$$\min_x \quad L(x, \tilde{\xi})$$
$$s.t. \quad g(x, \tilde{\xi}) \leq 0$$

Remarks:

- $\tilde{\xi}$ is unknown. Two main way of modelling it:
  - $\tilde{\xi} \in R$ with a known uncertainty set R, and a pessimistic approach. This is the robust optimization approach (RO).
  - $\tilde{\xi}$ is a random variable with known probability law. This is the Stochastic Programming approach (SP).
- Cost is not well defined.
  - RO : $\max_{\xi \in R} L(x, \xi)$.
  - SP : $\mathbb{E}\big[L(x, \boldsymbol{\xi})\big]$.
- Constraints are not well defined.
  - RO : $g(x, \xi) \leq 0, \qquad \forall \xi \in R$.
  - SP : $g(x, \boldsymbol{\xi}) \leq 0, \qquad \mathbb{P} - a.s.$.

An optimization problem with uncertainty

Adding uncertainty ξ in the mix

$$\min_x \quad L(x,\tilde{\xi})$$
$$s.t. \quad g(x,\tilde{\xi}) \leq 0$$

Remarks:

- $\tilde{\xi}$ is unknown. Two main way of modelling it:
  - $\tilde{\xi} \in R$ with a known uncertainty set R, and a pessimistic approach. This is the robust optimization approach (RO).
  - $\tilde{\xi}$ is a random variable with known probability law. This is the Stochastic Programming approach (SP).
- Cost is not well defined.
  - RO : $\max_{\xi \in R} L(x,\xi)$.
  - SP : $\mathbb{E}\big[L(x,\boldsymbol{\xi})\big]$.
- Constraints are not well defined.
  - RO : $g(x,\xi) \leq 0, \qquad \forall \xi \in R$.
  - SP : $g(x,\boldsymbol{\xi}) \leq 0, \qquad \mathbb{P}-a.s.$.

An optimization problem with uncertainty

Adding uncertainty ξ in the mix

$$\begin{aligned} \min_x \quad & L(x, \tilde{\xi}) \\ s.t. \quad & g(x, \tilde{\xi}) \leq 0 \end{aligned}$$

Remarks:

- $\tilde{\xi}$ is unknown. Two main way of modelling it:
  - $\tilde{\xi} \in R$ with a known uncertainty set R, and a pessimistic approach. This is the robust optimization approach (RO).
  - $\tilde{\xi}$ is a random variable with known probability law. This is the Stochastic Programming approach (SP).
- Cost is not well defined.
  - RO : $\max_{\xi \in R} L(x, \xi)$.
  - SP : $\mathbb{E}\big[L(x, \boldsymbol{\xi})\big]$.
- Constraints are not well defined.
  - RO : $g(x, \xi) \leq 0, \qquad \forall \xi \in R$.
  - SP : $g(x, \boldsymbol{\xi}) \leq 0, \qquad \mathbb{P} - a.s.$.

Requirements and limits

- Stochastic optimization :
  - requires a law of the uncertainty ξ
  - can be hard to solve (generally require discretizing the support and blowing up the dimension of the problem)
  - there exists specific methods (like Bender's decomposition)
- Robust optimization :
  - requires an uncertainty set R
  - can be overly conservative, even for reasonable R
  - complexity strongly depend on the choice of R
- Distributionally robust optimization :
  - is a mix between robust and stochastic optimization
  - consists in solving a stochastic optimization problem where the law is chosen in a robust way
  - is a fast growing fields with multiple recent results
  - but is still hard to implement than other approaches

Requirements and limits

- Stochastic optimization :
  - requires a law of the uncertainty ξ
  - can be hard to solve (generally require discretizing the support and blowing up the dimension of the problem)
  - there exists specific methods (like Bender's decomposition)
- Robust optimization :
  - requires an uncertainty set R
  - can be overly conservative, even for reasonable R
  - complexity strongly depend on the choice of R
- Distributionally robust optimization :
  - is a mix between robust and stochastic optimization
  - consists in solving a stochastic optimization problem where the law is chosen in a robust way
  - is a fast growing fields with multiple recent results
  - but is still hard to implement than other approaches

Requirements and limits

- Stochastic optimization :
  - requires a law of the uncertainty ξ
  - can be hard to solve (generally require discretizing the support and blowing up the dimension of the problem)
  - there exists specific methods (like Bender's decomposition)
- Robust optimization :
  - requires an uncertainty set R
  - can be overly conservative, even for reasonable R
  - complexity strongly depend on the choice of R
- Distributionally robust optimization :
  - is a mix between robust and stochastic optimization
  - consists in solving a stochastic optimization problem where the law is chosen in a robust way
  - is a fast growing fields with multiple recent results
  - but is still hard to implement than other approaches

Contents

Some numerical tests on real-life LPs

From Ben-Tal and Nemirovski

- take LP from Netlib library
- look at non-integer coefficients, assuming that they are not known with perfect certainty
- What happens if you change them by 0.1% ?
  - constraints can be violated by up to 450%
  - $\mathbb{P}(\text{violation} > 0) = 0.5$
  - $\mathbb{P}(\text{violation} > 150\%) = 0.18$
  - $\mathbb{E}[\text{violation}] = 125\%$

What do you want from robust optimization ?

- finding a solution that is less sensible to modified data, without a great increase of price
- choosing an uncertainty set R that:
  - offer robustness guarantee
  - yield an easily solved optimization problem

Contents

Solving a robust optimization problem

The robust optimization problem we want to solve is

$$\min_x \quad L(x)$$
$$s.t. \quad g(x,\xi) \leq 0 \qquad \forall \xi \in R$$

Note that, for simplicity reason we dropped w.l.o.g. the uncertainty in the objective.

Two main approaches are possible:

Constraint generation: replace R by a finite set of ξ, that is we replace an "infinite number of contraints" by a finite number of them.

Reformulation: replace $g(x,\xi) \leq 0 \quad \forall \xi \in R$, by $\sup_{\xi \in R} g(x,\xi) \leq 0$ and then explicit the sup.

Solving a robust optimization problem

The robust optimization problem we want to solve is

$$\begin{aligned} \min_x \quad & L(x) \\ s.t. \quad & g(x,\xi) \leq 0 \qquad\qquad \forall \xi \in R \end{aligned}$$

Note that, for simplicity reason we dropped w.l.o.g. the uncertainty in the objective.

Two main approaches are possible:

Constraint generation: replace R by a finite set of ξ, that is we replace an "infinite number of contraints" by a finite number of them.

Reformulation: replace $g(x,\xi) \leq 0 \quad \forall \xi \in R$, by $\sup_{\xi \in R} g(x,\xi) \leq 0$ and then explicit the sup.

Constraint generation

Data: Problem parameters, reference uncertainty ξ_0
Result: approximate value with gap;
for $k \in \mathbb{N}$ **do**
  solve $\tilde{v} = \min_x \{ L(x) \mid g(x, \xi_\kappa) \; \forall \kappa \leq k \}$ with optimal solution x_k;
  solve $s = \max_{\xi \in R} g(x_k, \xi)$ with optimal solution ξ_{k+1} ;
  if $s \leq 0$ **then**
    Robust optimization problem solved, with value $\tilde{v}$ and optimal solution x_k

Algorithm 1: Constraint Generation Algorithm

Note that we are solving a problem similar to the certain problem with an increasing number of constraints. This is easy to implement and can be numerically efficient.

Reformulation principle

We can write the robuste optimization problem as

$$\begin{aligned} \min_x \quad & L(x) \\ s.t. \quad & \sup_{\xi \in R} g(x,\xi) \leq 0 \end{aligned}$$

Now there are two way of simplifying this problem :

- either we can explicitely compute $\bar{g}(x) = \sup_{\xi \in R} g(x,\xi)$;
- or by duality we can write $\sup_{\xi \in R} g(x,\xi) = \min_{\zeta \in Q} h(x,\zeta)$
- $\min_{\zeta \in Q} h(x,\zeta) \leq 0$ is equivalent to $\exists \zeta$ such that $h(x,\zeta) \leq 0$, i.e. just add ζ as a variable in your optimization problem

Contents

Robust optimization for linear programm

We consider

$$\begin{aligned} \min_{x \geq 0} \quad & c^\top x \\ s.t. \quad & Ax \leq b \\ & \sup_{\xi \in R} \xi^\top x \leq b_i \qquad\qquad \forall i = 1..k \end{aligned}$$

On this example, for specific R, we are going to follow both reformulation approaches.

Contents

An explicit worst case value

- We consider an ellipsoidal uncertainty set

$$R = \Big\{\xi = \{a_i + \Delta_i u_i\}_i \quad | \quad \|u_i\|_2 \leq \rho\Big\}$$

- Here we can, for a given x, explicitely compute

$$\sup_{\xi \in R} \xi^\top x = a_i^\top x + \sup_{\|u_i\|_2 \leq \rho} (\Delta_i u_i)^\top x$$

$$= a_i^\top x + \rho\|\Delta_i x\|_2$$

- Hence, constraint

$$\sup_{\xi \in R} \xi^\top x \leq b_i$$

can be written

$$a_i^\top x + \rho\|\Delta_i x\|_2 \leq b_i$$

An explicit worst case value

- We consider an ellipsoidal uncertainty set

$$R = \Big\{ \xi = \{a_i + \Delta_i u_i\}_i \quad | \quad \|u_i\|_2 \leq \rho \Big\}$$

- Here we can, for a given x, explicitely compute

$$\begin{aligned} \sup_{\xi \in R} \xi^\top x &= a_i^\top x + \sup_{\|u_i\|_2 \leq \rho} (\Delta_i u_i)^\top x \\ &= a_i^\top x + \rho \|\Delta_i x\|_2 \end{aligned}$$

- Hence, constraint

$$\sup_{\xi \in R} \xi^\top x \leq b_i$$

can be written

$$a_i^\top x + \rho \|\Delta_i x\|_2 \leq b_i$$

An explicit worst case value

- We consider an ellipsoidal uncertainty set

$$R = \Big\{ \xi = \big\{ a_i + \Delta_i u_i \big\}_i \quad \Big| \quad \|u_i\|_2 \leq \rho \Big\}$$

- Here we can, for a given x, explicitely compute

$$\begin{aligned} \sup_{\xi \in R} \xi^\top x &= a_i^\top x + \sup_{\|u_i\|_2 \leq \rho} (\Delta_i u_i)^\top x \\ &= a_i^\top x + \rho \|\Delta_i x\|_2 \end{aligned}$$

- Hence, constraint

$$\sup_{\xi \in R} \xi^\top x \leq b_i$$

can be written

$$a_i^\top x + \rho \|\Delta_i x\|_2 \leq b_i$$

SOCP problem

- An Second Order Cone Programming constraint is a constraint of the form

$$\|a^\top x + b\|_2 \leq c^\top x + d$$

- An SOCP problem is a (continuous) optimization problem with linear cost and linear and SOCP constraints
- There exists powerful software to solve SOCP (e.g. CPLEX, Gurobi, MOSEK...) with dedicated interior points methods
- There exist a duality theory akin to the LP duality theory
- If a robust optimization problem can be cast as an SOCP the formulation is deemed efficient

Contents

Linear duality : recalls

- Recall that, if finite,

$$\begin{aligned} \max_{\xi} \quad & \xi^\top x \\ s.t. \quad & D\xi \le d \end{aligned}$$

as the same value as

$$\begin{aligned} \min_{\zeta} \quad & \zeta^\top d \\ s.t. \quad & \zeta^\top D = x \\ & \zeta \ge 0 \end{aligned}$$

- Thus,

$$\begin{aligned} \sup_{\xi : D\xi \le d} \quad \xi^\top x \le b \quad & \iff \quad \min_{\zeta \ge 0 : \zeta^\top D = x} \quad \zeta^\top d \le b \\ & \iff \quad \exists \zeta \ge 0, \quad \zeta^\top D = x, \quad \zeta^\top d \le b \end{aligned}$$

Linear duality : recalls

- Recall that, if finite,

$$\begin{aligned} \max_{\xi} \quad & \xi^\top x \\ s.t. \quad & D\xi \leq d \end{aligned}$$

as the same value as

$$\begin{aligned} \min_{\zeta} \quad & \zeta^\top d \\ s.t. \quad & \zeta^\top D = x \\ & \zeta \geq 0 \end{aligned}$$

- Thus,

$$\begin{aligned} \sup_{\xi : D\xi \leq d} \xi^\top x \leq b \quad & \iff \quad \min_{\zeta \geq 0 : \zeta^\top D = x} \zeta^\top d \leq b \\ & \iff \quad \exists \zeta \geq 0, \quad \zeta^\top D = x, \quad \zeta^\top d \leq b \end{aligned}$$

Polyhedral uncertainty

- We consider a polyhedral uncertainty set

$$R = \left\{ \xi \quad | \quad D\xi \leq d \right\}$$

- Then the robust optimization problem

$$\begin{aligned} \min_{x \geq 0} \quad & c^\top x \\ s.t. \quad & \sup_{\xi \in R} \xi^\top x \leq h \end{aligned}$$

reads

$$\begin{aligned} \min_{x \geq 0, \zeta \geq 0} \quad & c^\top x \\ s.t. \quad & \zeta^\top d \leq h \\ & \zeta^\top d = x \end{aligned}$$

Contents

Soyster model

The problem

$$\begin{aligned} \min_{x} \quad & c^\top x \\ & \sup_{\tilde{A} \in R} \tilde{A}x \leq b \\ & \underline{x} \leq x \leq \bar{x} \end{aligned}$$

where each coefficient $\tilde{a}_{ij} \in [\bar{a}_{ij} - \delta_{ij}, \bar{a}_{ij} + \delta_{ij}]$ can be written

$$\begin{aligned} \min_{x} \quad & c^\top x \\ & \sum_j \bar{a}_{ij} x_j + \sum_j \delta_{ij} |x_j| \leq b_i \qquad \forall i \\ & \underline{x} \leq x \leq \bar{x} \end{aligned}$$

Soyster model

The problem

$$\min_x \quad c^\top x$$
$$\sup_{\tilde{A} \in R} \tilde{A}x \leq b$$
$$\underline{x} \leq x \leq \bar{x}$$

where each coefficient $\tilde{a}_{ij} \in [\bar{a}_{ij} - \delta_{ij}, \bar{a}_{ij} + \delta_{ij}]$ can be written

$$\min_{x,y} \quad c^\top x$$
$$\sum_j \bar{a}_{ij} x_j + \sum_j \delta_{ij} y_j \leq b_i \qquad \forall i$$
$$\underline{x} \leq x \leq \bar{x}$$
$$y_j \geq x_j, \quad y_j \geq -x_j$$

Cardinality constrained LP I

Soyster's model is over conservative, we want to consider a model where only Γ_i coefficient per line have non-zero errors, leading to

$$\begin{aligned} \min_{x,y} \quad & c^\top x \\ & \sum_j \bar{a}_{ij} x_j + \max_{S_i : |S_i| = \Gamma_i} \sum_{j \in S_i} \delta_{ij} y_j \leq b_i \qquad \forall i \\ & \underline{x} \leq x \leq \bar{x} \\ & y_j \geq x_j, \quad y_j \geq -x_j \end{aligned}$$

Cardinality constrained LP II

This means that, for line i we take a margin of

$$\beta_i(x,\Gamma_i) := \max_{S_i:|S_i|=\Gamma_i} \sum_{j\in S_i} \delta_{ij}|x_j|$$

which can be obtained as

$$\begin{aligned} \beta_i(x,\Gamma_i) = \max_{z} \quad & \sum_j \delta_{ij}|x_j| z_{ij} \\ & \sum_j z_{ij} \leq \Gamma_i \\ & 0 \leq z_{ij} \leq 1 \end{aligned}$$

This LP can be then dualized to be integrated in the original LP.

Cardinality constrained LP III

In the end we obtain

$$
\begin{aligned}
\min_{x,\beta,\lambda,\mu} \quad & c^\top x \\
& \sum_j \bar{a}_{ij} x_j + \beta_i \leq b_i & \forall i \\
& \lambda_i \Gamma_i + \sum_j \mu_{ij} \leq \beta_i & \forall i \\
& \delta_{ij} x_j \leq \lambda_i + \mu_{ij} & \forall i,j \\
& -\delta_{ij} x_j \leq \lambda_i + \mu_{ij} & \forall i,j \\
& \lambda \geq 0, \quad \mu \geq 0 \\
& \underline{x} \leq x \leq \bar{x}
\end{aligned}
$$

Contents

A combinatorial optimization problem with cardinality constraint

We consider a combinatorial optimization problem:

$$\min_{x\in\{0,1\}^N} \quad \max_{\tilde{c}\in R} \tilde{c}^\top x$$
$$s.t. \quad x \in X$$

where R is such that each $\tilde{c}_i \in [\bar{c}_i, \bar{c}_i + \delta_i]$, with at most Γ coefficient deviating from $\bar{c}_i$.

Thus the problem reads

$$(P) \qquad \min_{x\in\{0,1\}^N} \quad \bar{c}^\top x + \max_{|S|\leq\Gamma} \sum_{i\in S} \delta_i x_i$$
$$s.t. \quad x \in X$$

wlog we assume that the i are ordered by decreasing cost uncertainty span : $\delta_1 \geq \delta_2 \geq \cdots \geq \delta_n$.

A combinatorial optimization problem with cardinality constraint

We consider a combinatorial optimization problem:

$$\min_{x\in\{0,1\}^N} \quad \max_{\tilde{c}\in R} \tilde{c}^\top x$$
$$s.t. \quad x \in X$$

where R is such that each $\tilde{c}_i \in [\bar{c}_i, \bar{c}_i + \delta_i]$, with at most Γ coefficient deviating from $\bar{c}_i$.
Thus the problem reads

$$(P) \qquad \min_{x\in\{0,1\}^N} \quad \bar{c}^\top x + \max_{|S|\leq\Gamma} \sum_{i\in S} \delta_i x_i$$
$$s.t. \quad x \in X$$

wlog we assume that the i are ordered by decreasing cost uncertainty span : $\delta_1 \geq \delta_2 \geq \cdots \geq \delta_n$.

Solving the robust combinatorial problem I

We can write (P) as

$$\begin{aligned} \min_{x\in\{0,1\}^N} \max_{u\in[0,1]^n} \quad & \bar{c}^\top x + \sum_{i=1}^n \delta_i x_i u_i \\ s.t. \quad & x \in X \\ & \sum_{i=1}^n u_i \leq \Gamma \end{aligned}$$

For a given $x \in X$ we dualize the inner maximization LP problem

Solving the robust combinatorial problem II

Thus we can write (P) as

$$\begin{aligned} \min_{x,y,\theta} \quad & \bar{c}^\top x + \Gamma\theta + \sum_{j=1}^{n} y_j \\ s.t. \quad & x \in X \\ & y_j + \theta \geq \delta_j x_j \\ & y_j, \theta \geq 0 \end{aligned}$$

Note that an optimal solution satisfies

$$y_j = (\delta_j x_j - \theta)^+ = (\delta_j - \theta)^+ x_j$$

as $x_j \in \{0,1\}$, and $\theta \geq 0$.

Solving the robust combinatorial problem II

Thus we can write (P) as

$$\begin{aligned} \min_{x,y,\theta} \quad & \bar{c}^\top x + \Gamma\theta + \sum_{j=1}^{n} y_j \\ s.t. \quad & x \in X \\ & y_j + \theta \geq \delta_j x_j \\ & y_j, \theta \geq 0 \end{aligned}$$

Note that an optimal solution satisfies

$$y_j = (\delta_j x_j - \theta)^+ = (\delta_j - \theta)^+ x_j$$

as $x_j \in \{0, 1\}$, and $\theta \geq 0$.

Solving the robust combinatorial problem III

Thus we can write (P) as

$$\min_{x,\theta \geq 0} \quad \bar{c}^\top x + \Gamma\theta + \sum_{j=1}^{n} x_j(\delta_j - \theta)^+$$

$$s.t. \quad x \in X$$

We can now decompose the problem for $\theta \in [\delta_\ell, \delta_{\ell-1}]$ where $\delta_{n+1} = 0$ and $\delta_0 = +\infty$.
Therefore, we have

$$val(P) = \min_{\ell \in [n]} Z^\ell$$

where

$$Z^\ell = \min_{x \in X, \theta \in [\delta_\ell, \delta_{\ell-1}]} \quad \bar{c}^\top x + \Gamma\theta + \sum_{j=1}^{\ell-1} x_j(\delta_j - \theta)$$

Solving the robust combinatorial problem IV

As the problem is linear in θ we have that

$$Z^\ell = \min_{x \in X, \theta \in [\delta_\ell, \delta_{\ell-1}]} \quad \bar{c}^\top x + \Gamma\theta + \sum_{j=1}^{\ell-1} x_j(\delta_j - \theta)$$

is attained for $\theta = \delta_\ell$ or $\theta = \delta_{\ell-1}$.
So in the end, we have

$$val(P) = \min_{\ell \in [n]} G^\ell$$

where

$$G^\ell = \Gamma\delta_\ell + \min_{x \in X} \left\{ \bar{c}^\top x + \sum_{i=1}^{\ell} \underbrace{(\delta_i - \delta_\ell)}_{\geq 0} x_j \right\}$$

Algorithm for the robust problem

1. For $\ell \in [n]$, solve

$$G^\ell = \Gamma\delta_\ell + \min_{x\in X} \quad \Big\{\bar{c}^\top x + \sum_{i=1}^{\ell}(\delta_i - \delta_\ell)x_j\Big\}$$

   with optimal solution x_ℓ

2. Set $\ell^* \in \arg\min_{\ell\in[n]} G^\ell$
3. Return $val(P) = G^{\ell^*}$ and $x^* = x_\ell$

Contents

Why do robust optimization ?

- Because you want to account for some uncertainty
- Because you want to have a solution that resists to changes in data
- Because your data is unprecise and robustness yield better out-of-sample result
- Because you do not have the law of the uncertainty
- Because you can control the robustness level
- Because your problem is "one-shot"

Which uncertainty set to choose ?

- An uncertainty set that is computationally tractable
- An uncertainty set that yields good results
- An uncertainty set that have some theoretical soundness
- An uncertainty set that take available data into account
- Select uncertainty set / level through cross-validation

Is there some theoretical results ?

- Yes: with some assumption over the randomness (e.g. bounded and symmetric around $\bar{a}$) some uncertainty set (e.g. ellipsoidal) have a probabilistic guarantee :

$$\forall \xi \in R_\varepsilon, \quad g(x, \xi) \leq 0 \qquad \Longrightarrow \qquad \mathbb{P}\Big(g(x, \boldsymbol{\xi}) \leq 0\Big) \geq 1 - \varepsilon$$

- Yes: in some cases approximation scheme for nominal problem can be extended to robust problem (e.g. cardinal uncertainty in combinatorial problem)
- Yes: using relevant data we can use statistical tools to construct a robust set R that imply a probabilistic guarantee

D. Bertsimas, D. Brown, C. Caramais
Theory and applications of robust optimization
Siam Review, 2011.

BL Gorissen, I. Yanikoglu and D. den Hertog
A practical guide to robust optimization
Omega, 2015.

D. Bertsimas and M. Sim
The price of robustness
Operations research, 2004.

D. Bertsimas and M. Sim
Robust discrete optimization and network flows
Mathematical Programming, 2003.

A. Ben Tal, L El Ghaoui, A. Nemirovski
Robust optimization
Springer, 2009.