An introduction to Stochastic Optimisation

Emilien Flayac & Alejandro I. Maass

The University of Melbourne

Broad view on optimisation

Optimisation is about making a choice among a range of possible options according to some quantitative criterion

An optimisation problem is defined by:

- A *feasible* (or *admissible*) set $\mathcal{X}$ that represents the range of our possibilities
- A *cost* (or *objective*) function $f : \mathcal{X} \to \mathbb{R}$ that weighs every feasible option

$$\min_{x \in \mathcal{X}} f(x) \tag{P}$$

Solution

x^* is a solution of (P) iff $x^* \in \mathcal{X}$ and $\forall x \in \mathcal{X}$, $f(x^*) \leq f(x)$

Broad view on optimisation: Admissible set

The nature of $\mathcal{X}$ depends on the nature of the decision:

- Discrete vs. Continuous
- Finite vs. Infinite dimensional
- Static vs. Dynamic
- Defined by a finite set of inequality constraints:

$$\mathcal{X} = \{x \in E : c(x) \leq 0\}$$

with $c(x) = (c_1(x), \ldots, c_m(x))$

Examples of constraints:

- *physical restrictions*: positiveness, equilibrium laws, ...
- *hard goals*: target points, production objectives, ...
- *safety issues*: "staying on the road", collision avoidance, ...

Broad view on optimisation: Cost function

f allows one to compare the available choices from $\mathcal{X}$.

The nature of f (and c) defines several types of problems:

- Linear vs. Nonlinear
- Differentiable vs. Nondifferentiable
- Convex vs. Nonconvex

Examples of objective functions:

- *actual cost to minimise*: fuel consumption, expenses, ...
- *profit to maximise*: return on investment, sales, ...
- *other performance indices*: time, distance, velocity, ...

Typical problem:

$$\min_{c(x) \leq 0} f(x) \qquad (P_c)$$

Uncertainty in optimisation: Sources

In some cases, unpredictable events occur during the optimisation process

Examples of sources of uncertainty:

- *Power systems, Inventory management*: uncertainty on the demand
- *Portfolio selection*: risks in the share market
- *Biological systems*: climate, predation
- *System stabilization*: modelling error, noise

Uncertainty in optimisation: perturbed cost and constraints

Unknown parameter belonging to an uncertainty set:

$$\boxed{d \in \mathcal{D}}$$

It acts on:

- The cost $\Rightarrow F(x, d)$
- The constraints $\Rightarrow C(x, d)$

Main issues:

- Can we make d less uncertain ?
  Get information (observe) if we can !
- If not, how to write a proper uncertain optimisation problem ?
  Need to model the uncertainty !

Observation first vs Decision first

When one is able to observe d then two cases are possible:

- The observation comes *before* the decision (Observation first) $\Rightarrow d$ is known and we get a parametrized optimization problem:

$$\min_{c_d(x)\leq 0} f_d(x) \tag{P_d}$$

  with $c_d(x) = C(x,d)$, $f_d(x) = F(x,d)$ and solutions x_d^*

- The observation comes *after* the decision (Decision first) $\Rightarrow d$ remains unknown and the problem depends on the representation of the uncertainty

Pitfall to avoid

Consider one value d_0 and work with this one only

Worst case approach

Idea: consider every situation identically and make a decision based on assuming the worst possible value of d

- Cost: $f(x) = \max_{d \in \mathcal{D}} F(x,d)$
  As if 'nature' is playing against us and we have to assume its best response
- Admissible set: $\mathcal{X} = \{x : \max_{d \in \mathcal{D}} C(x,d) \leq 0\}$
  The constraint must be satisfied for any value of d
  $\Rightarrow$ *robust* formulation

Remarks:

- The uncertainty is here characterized by $\mathcal{D}$
- Very safe but very restrictive so usually not very efficient
  $\Rightarrow$ *conservative* approach

How to avoid conservativeness? Stochastic approach

Stochastic optimisation

General Idea: consider the various situations differently based on their probability of occurrence

$\Rightarrow$ Represent the uncertainty as a random variable D valued in $\mathcal{D}$

Stochastic optimisation: Expected cost

Idea: make a decision based on the average cost, on a aggregation of the scenarios

- Cost: $f(x) = \mathbb{E}[F(x, D)]$
- Admissible set: $\mathcal{X} = \{x : \; c(x) \leq 0\}$

Remarks:

- The uncertainty is characterized here by the probability distribution of D
- Typically less conservative but less safe as one does not care about particular realizations of D
- f keeps many properties of F (linearity, convexity, differentiability,. . .)

Stochastic optimisation: Constraint in expectation

Idea: put a restriction on the average value of the constraint

- Cost: $f(x)$
- Admissible set: $\mathcal{X} = \{x : \; c(x) := \mathbb{E}[C(x, D)] \leq 0\}$

Remarks:

- safer than just a expected cost but not very intuitive
- c keeps many properties of C (linearity, convexity, differentiability,...)

Stochastic optimisation: Chance constraint

Idea: put a constraint on the probability of admissibility

- Cost: $f(x)$
- Admissible set: $\mathcal{X} = \{x : \mathbb{P}(C(x, D) \leq 0) \geq 1 - \alpha\}$
  with $\alpha > 0$ being the level of risk

Remarks:

- Stochastic equivalent of the robust formulation
- $\mathcal{X}$ becomes a complicated set

Problems with recourse

What if we could make two decisions, one before observing D and one after ?

- Choose $x_0 \in \mathcal{X}_0$ and pay $f_0(x_0)$
- Observe D
- Choose $X_1 = \pi_1(D) \in \mathcal{X}_1$ and pay $f_1(x_0, X_1, D)$

We get the following two-stage problem:

$$\min_{x_0 \in \mathcal{X}_0, X_1 = \pi_1(D) \in \mathcal{X}_1,} \mathbb{E}[f_0(x_0) + f_1(x_0, X_1, D)]$$

Remarks:

- Typically infinite dimensional problem even if $\mathcal{X}_0$ and $\mathcal{X}_1$ are of finite dimension
- Usually solved by backward decomposition
- Can be extended to multi-stage problem

Numerical methods

General remarks

- One can combine the formulations !
- They are versions of (P_c) with particular structures where f and c are hard to compute

Examples of numerical methods

- *Expected cost*: Stochastic average approximation (SAA), Stochastic gradient methods, etc ...
- *Constraint in expectation*: dual methods, primal-dual methods, etc ...
- *Chance constraint*: deterministic reformulation or approximation, scenario approach, etc ...
- *Multi-stage problems*: Dynamic Programming, Scenario trees, Reinforcement learning, etc ...

Inventory: The newsvendor problem

- A company decides about ordering quantity x of a certain product to satisfy demand d (unknown at purchase time)
- Cost per unit is c
- If $d > x$, additional order at unit price $b \geq 0$ is made
- $b > c$, i.e. backorder penalty cost is larger than ordering
- If $d < x$, a holding cost of $h(d - x)$ is incurred

Total cost

$$\boxed{F(x, d) = cx + b \max\{d - x, 0\} + h \max\{x - d, 0\}}$$

Objective

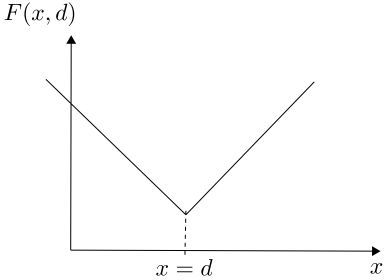
Minimise the total cost $F(x, d)$, where x is the decision variable and d is a parameter.

Cost function

We can re-write the objective function as

$$F(x, d) = \max\{(c - b)x + bd, (c + h)x - hd\},$$

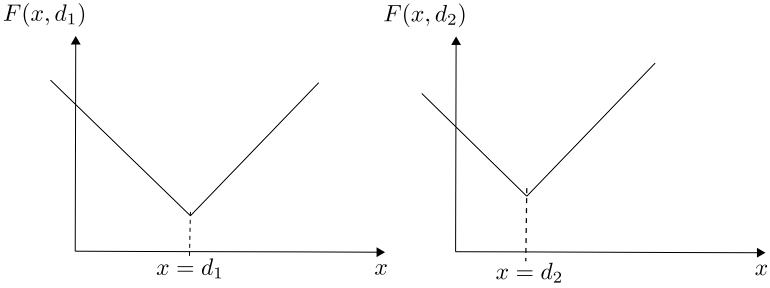
which is a piecewise linear function with minimum attained at $\bar{x} = d$, provided the demand is **known!**



What should we do when d is unknown when ordering?

The ordering decision should be made **before** a realisation of the demand becomes known. View demand D as a random variable (d is a particular realisation of D)

- **Silly attempt:** Optimise $\min_{x \geq 0} F(x, D)$
- This problem does not make sense! $F(x, D)$ is random



- One x cannot simultaneously optimise both functions!

Attempt 2: Plan for the average case

- Suppose the cdf of D is given by $H(x) := \mathbb{P}(D \leq x)$
- Let $\mu := \mathbb{E}[D]$
- Then, $\arg\min_{x \geq 0} F(x, \mu) = \mu$
- The optimal policy is to order a quantity μ

Remark

This approach might be more sensitive to random perturbations of the empirical data.

Attempt 3: Plan for worst case

- Suppose that we know upper and lower bounds on the demand d, i.e. $\ell \leq d \leq u$
- Consider $\min_{x \geq 0} \{ f(x) := \max_{d \in [\ell, u]} F(x, d) \}$
- While making a decision x, one is prepared for the worst possible outcome of the maximal cost
- The optimal solution is

$$x^* = \frac{h\ell + bu}{h + b}$$

- If the holding cost is zero ($h = 0$), then the optimal solution is obviously $x^* = u$ (always order the maximum demand)
- Often a conservative approach!

Proof

- Recall that $F(x,d) = \max\{(c-b)x + bd, (c+h)x - hd\}$, then $\max_{d\in[\ell,u]} F(x,d) = \max\{F(x,\ell), F(x,u)\}$
- We should look at the problem in $x \in [\ell, u]$, then

$$
\begin{aligned}
&\min_{x\geq 0} \max_{d\in[\ell,u]} F(x,d) \\
&\qquad\qquad = \min_{x\in[\ell,u]} \max\{F(x,\ell), F(x,u)\} \\
&\qquad\qquad = \min_{x\in[\ell,u]} \max\{cx + h[x-\ell]_+, cx + b[u-x]_+\}
\end{aligned}
$$

- This is a piecewise linear convex function, and the optimal solution is attained at the point where $h(x-\ell) = b(u-x)$, completing the proof.

A suitable framework

- The probability distribution of D is known
- It makes sense to talk about the expected value $\mathbb{E}\left[F(x,D)\right]$ of the total cost as a function of x
- Consequently,

$$\min_{x\geq 0}\{f(x) := \mathbb{E}\left[F(x,D)\right]\}$$

- We optimise the total cost **on average**

Solution to $\min_{x \geq 0} \mathbb{E}\left[F(x, D)\right]$

- This particular example can be solved in closed form
- Recall $H(x) \coloneqq \mathbb{P}\{D \leq x\}$, $H(x) = 0$ for all $x < 0$
- The optimal solution is equal to the quantile

$$x^* = H^{-1}\left(\frac{b-c}{b+h}\right)$$

It is not always this easy...

- The newsvendor problem is about the only stochastic program that admits a simple "closed-form" solution
- In general, we must solve instances **numerically**!

Proof

- Recall that

$$\begin{aligned} F(x, D) &= cx + b\max\{D - x, 0\} + h\max\{x - D, 0\} \\ &= \max\{(c - b)x + bD, (c + h)x - hD\} \end{aligned}$$

- $f(x) := \mathbb{E}[F(x, D)]$ is convex and continuous
- For $x \geq 0$, $f(x) = f(0) + \int_0^x f'(z)dz$. Note that $f(0) = b\mathbb{E}[D]$ since $D \geq 0$
- Moreover,

$$\begin{aligned} f'(z) &= \frac{\partial}{\partial z}\mathbb{E}\left[cz + b\max\{D - z, 0\} + h\max\{z - D, 0\}\right] \\ &= c - b\mathbb{P}[D \geq z] + h\mathbb{P}[D \leq z] \\ &= c - b(1 - H(z)) + hH(z) \\ &= c - b + (b + h)H(z) \end{aligned}$$

Proof

Then,

$$\mathbb{E}[F(x,D)] = b\mathbb{E}[D] + (c-b)x + (b+h)\int_0^x H(z)dz$$

To obtain the minimum we take the derivative of the right-hand side and equate it to zero. Consequently, the optimal solutions satisfy

$$(b+h)H(x^*) + c - b = 0$$

and the result follows.

Numerical comparison

$$F(x, D) = cx + b[D - x]_+ + h[x - D]_+$$

- **Cost per unit:** $c = 1$
- **Re-ordering cost:** $b = 5$
- **Holding cost:** $h = 0.5$
- **Demand:** Normally distributed $\mathcal{N}(100, 14^2)$

Plan for	Optimal solution x^*	$\mathbb{E}[F(x^*, D)]$
Average demand	100	\$ 131
Worst case ($58 \leq D \leq 142$)	134	\$ 152
Expected cost	112	\$ 126

Chance constraints

Remark

For particular realisations of the demand, $F(x^*, D)$ can be quite different from $\mathbb{E}[F(x^*, D)]$. A natural question is whether we can control the risk of the cost $F(x, D)$ to be not "too high".

- For a given threshold $\tau > 0$, we want $F(x, D) \leq \tau$ to be satisfied for all possible realisations of $D \in \mathcal{D}$
- Can be quite restrictive if the uncertainty set $\mathcal{D}$ is large!
- We thus introduce a constraint

$$\mathbb{P}\{F(x, D) \leq \tau\} \geq 1 - \alpha$$

- For $x \leq \tau/c$, this becomes

$$H\left(\frac{(b-c)x+\tau}{b}\right) - H\left(\frac{(c+h)x-\tau}{h}\right) \geq 1 - \alpha$$

Newsvendor problem

The newsvendor is a simple example of a **two-stage** problem

- **First stage:** Before a realisation of D is known, one makes a decision about x
- **Second stage:**
  - After a realisation d of D becomes known, it may happen that $d > x$ (Random "stuff" happens)
  - The company then takes a recourse action of ordering $d - x$ at a higher cost of $b > c$ (We attempt to repair the havoc)

The evolution of information is of paramount importance!

THANK YOU FOR YOUR ATTENTION!