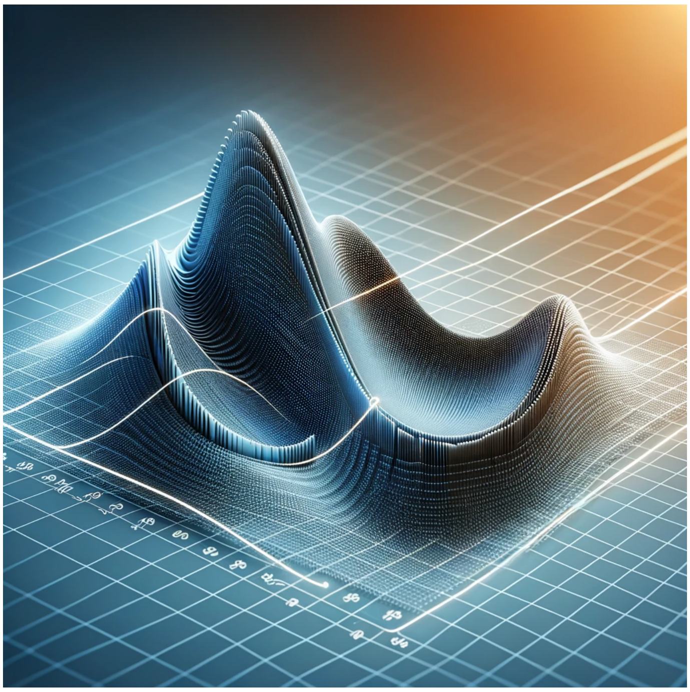


LINMA2491 Operational Research

Simon Desmidt
Issambre L'Hermite Dumont

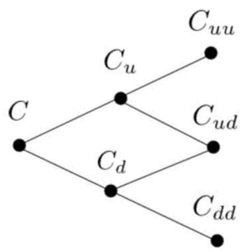
Academic year 2024-2025 - Q2

UCLouvain

Contents

Definition and notation

- Given Ω, a sigma-algebra $\mathcal{A}$ is a set of subsets of Ω, with the elements called events, such that:
  - $\Omega \in \mathcal{A}$
  - if $A \in \mathcal{A}$ then also $\Omega - A \in \mathcal{A}$
  - if $A_i \in \mathcal{A}$ for $i = 1, 2, \ldots$ then also $\cup_{i=1}^{\infty} A_i \in \mathcal{A}$
  - if $A_i \in \mathcal{A}$ for $i = 1, 2, \ldots$ then also $\cap_{i=1}^{\infty} A_i \in \mathcal{A}$
- Consider:



  - The state space is the set of all values of the system at each stage.

$$S_0 = \{C\}, \qquad S_1 = \{C_u, C_d\}, \qquad S_2 = \{C_{uu}, C_{ud}, C_{dd}\} \tag{1.1}$$

  - The sample space is the set of all possible combination of the system.

$$\Omega = S_0 \times S_1 \times S_2 = \{(C, C_u, C_{uu}), (C, C_u, C_{ud}), (C, C_u, C_{dd}), \ldots\} \tag{1.2}$$

- The power set of Ω is the set of all of the subsets, denoted $\mathcal{B}(\Omega)$.
- The probability space is the triplet $(\Omega, \mathcal{A}, P)$ where P is a probability measure.
  - $P(\emptyset) = 0$
  - $P(\Omega) = 1$
  - $P(\cup_{i=1}^{\infty} A_i) = \sum_i P(A_i)$ if A_i are disjoint
- $\forall t$, A_t is the set of events on which we have information at stage t. For example, $A_0 = \{C\}$, $A_1 = \{C, C_u, C_d\}$. Thus is it evident that $t_1 \leq t_2 \Rightarrow \mathcal{A}_{t_1} \subseteq \mathcal{A}_{t_2}$

- Consider the following problem with $x \in \mathbb{R}^n$ and domain $\mathcal{D}$:

$$\begin{aligned} &\min f_0(x), \qquad s.t. \\ &f_i(x) \leq 0, i = 1, \dots, m \\ &h_j(x) = 0, j = 1, \dots, p \end{aligned} \tag{1.3}$$

  Then the Lagrangian function is defined as $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$:

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{p} \nu_j h_j(x) \tag{1.4}$$

- The Lagrange dual function is defined as $g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$:

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \tag{1.5}$$

- The Lagrange dual problem is a lower bound on the optimal value of the primal problem

- Lagrange relaxation of Stochastic Programs, consider the two problems:

$$\begin{aligned} &\min f_1(x) + \mathbb{E}_\omega[f_2(y(\omega), \omega)] & &\min f_1(x) + \mathbb{E}_\omega[f_2(y(\omega), \omega)] \\ &s.t \qquad h_{1i}(x) \leq 0, i = 1, \dots, m_1 & &s.t. \qquad h_{1i}(x) \leq 0, i = 1, \dots, m_1 \\ &h_{2i}(x, y(\omega), \omega) \leq 0, i = 1, \dots, m_2 & &h_{2i}(x(\omega), y(\omega), \omega) \leq 0, i = 1, \dots, m_2 \\ & & &\color{red}{x(\omega) = x} \end{aligned} \tag{1.6}$$

  The red constraint is the non-anticipativity constraint, it transforms the deterministic variable into a stochastic variable. A VERIFIER

- The dual of a stochastic program is:

$$\begin{aligned} &g(\nu) = g1(\nu) + \mathbb{E}_\omega(g2(\nu, \omega)) \\ \textit{where} \\ &g_1(\nu) = \inf f_1(x) + \left(\sum_{\omega \in \Omega} \nu(\omega) \right)^T x \\ &s.t. \quad h_{1i}(x) \leq 0, i = 1, \dots, m_1 \\ \textit{and} \\ &g_2(\nu, \omega) = \inf f_2(y(\omega)\omega) - \nu x(\omega) \\ &s.t. \quad h_{2i}(x(\omega), y(\omega), \omega) \leq 0, i = 1, \dots, m_2 \end{aligned} \tag{1.7}$$

- With p^* the solution of the primal problem and d^* the solution of the dual problem, we have:
  - Weak duality: $d^* \leq p^*$
  - Strong duality: $d^* = p^*$

- The KKT conditions are necessary and sufficient for optimality in convex optimization, there aren't unique. They are:

- Primal constraint: $f_i(x) \leq 0, i = 1, \ldots, m, h_j(x) = 0, j = 1, \ldots, p$
- Dual constraint: $\lambda \geq 0$
- Complemetarity slackness: $\lambda_i f_i(x) = 0, i = 1, \ldots, m$
- Gradient of the Lagrangian: $\nabla_x L(x, \lambda, \nu) = 0$

- An extreme point of a polyhedron P is a point $x \in P$ such that it cannot be expressed as a linear combination of two distinct points in P, i.e. an extreme point is a vertex of the polyhedron.

- An extreme ray of a polyhedron P is $\sigma \in \mathbb{R}^n$ such that for all $x \in P$, for all $\lambda \in [0, 1]$,

$$(x + \lambda\sigma) \in P \tag{1.8}$$

  i.e. it is a direction in which we can travel infinitely without leaving the polyhedron.

1.1 Reminders on subgradients

π is a subgradient of the function g at u if

$$g(w) \geq g(u) + \pi^T(w - u) \qquad \forall w \tag{1.9}$$

If $g = \max\{g_1, g_2\}$ with $g_{1,2}$ convex and differentiable, the subgradient of g at u_0 is

- $\pi = \nabla g_1(u_0)$ if $g_1(u_0) > g_2(u_0)$
- $\pi = \nabla g_2(u_0)$ if $g_2(u_0) > g_1(u_0)$
- The line segment $[\nabla g_1(u_0), \nabla g_2(u_0)]$ if $g_1(u_0) = g_2(u_0)$

The subdifferential of g at u is the set of all subgradients of g at u, denoted $\partial g(u)$. If g is convex, then its subdifferential is nonempty on its domain, and g is differentiable at u if its $\partial g(u) = \{\pi\}$.

1.1.1 Use in duality

Define $c(u)$ as the optimal value of

$$\begin{aligned} c(u) &= \min f_0(x) \\ f_i(x) &\leq u_i \qquad i = 1, \ldots, m \end{aligned} \tag{1.10}$$

where $x \in dom f_0$ and f_0, f_i are convex functions.

- $c(u)$ is convex;

- If strong duality holds, denote $\lambda^\star$ as the maximizer of the dual function

$$\inf_{x \in dom f_0} (f_0(x) - \lambda^T(f(x) - u)) \tag{1.11}$$

  for $\lambda \leq 0$. Then, $\lambda^\star \in \partial c(u)$. λ_i represents the sensitivity of $c(u)$ to a marginal change in the right-hand side of the i-th constraint.

Modelling

2.1 Introduction

- For a certain sequence of events $x \to \omega \to y(\omega)$, where ω is the uncertainty,
  - A first-stage decision is a decision that is made before the uncertainty is revealed (i.e. in x);
  - A second-stage decision is a decision that is made after the uncertainty is revealed (i.e. in $y(\omega)$).
- We can have the following mathematical formulation:

$$\begin{aligned} \min c^T x + \mathbb{E}[\min q(\omega)^T y(\omega)] \\ Ax = b \\ T(\omega)x + W(\omega)y(\omega) = h(\omega) \\ x \geq 0, y(\omega) \geq 0 \end{aligned} \tag{2.1}$$

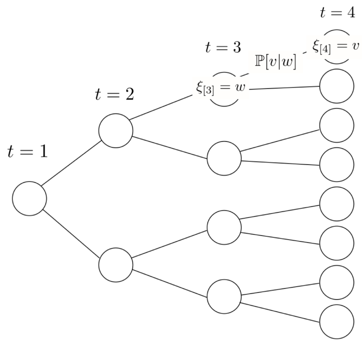
  - First-stage decision variable: $x \in \mathbb{R}^{n_1}$
  - First-stage parameter: $c \in \mathbb{R}^{n_1}$, $b \in \mathbb{R}^{m_1}$ and $A \in \mathbb{R}^{m_1 \times n_1}$
  - Second-stage decision: $y(\omega) \in \mathbb{R}^{n_2}$
  - Second-stage data: $q(\omega) \in \mathbb{R}^{n_2}$, $h(\omega) \in \mathbb{R}^{m_2}$ and $T(\omega) \in \mathbb{R}^{m_2 \times n_1}$, $W(\omega) \in \mathbb{R}^{m_2 \times n_2}$

2.2 Representations

2.2.1 Scenario Trees

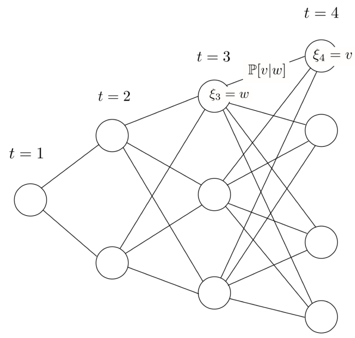
A scenario tree is a graphical representation of a Markov process $\{\xi_t\}_{t \in \mathbb{Z}}$, where the nodes are the history of realizations ($\xi_{[t]} = (\xi_1, \ldots, \xi_t)$), and the edges are the transitions from $\xi_{[t]}$ to $\xi_{[t+1]}$.

- We denote the root as $t = 1$;
- An ancestor of a node $\xi_{[t]}$, $A(\xi_{[t]})$ is a unique adjacent node which precedes ξ_t;
- The children of a node, $C(\xi_{[t]})$ are the nodes that are adjacent to $\xi_{[t]}$ and occur at stage $t + 1$.



2.2.2 Lattice

A lattice is a graphical representation of a Markov process $\{\xi_t\}_{t\in\mathbb{Z}}$, where the nodes are the realizations ξ_t and the edges correspond to the transitions from ξ_t to ξ_{t+1}.



2.2.3 Serial Independence

A process satisfies serial independence if, for every stage t, ξ_t has a probability distribution that does not depend on the history of the process. Thus, the probability measure is

$$\mathbb{P}\left[\xi_t(\omega)=i\middle|\xi_{[t-1]}(\omega)\right]=p_t(i)\qquad\forall\,\xi_{[t-1]}\in\Xi_{[t-1]},i\in\Xi_t\tag{2.2}$$

2.3 Multi Stage Stochastic Linear Program

2.3.1 Notation

- Probability space: $(\Omega,2^\Omega,\mathbb{P})$ with filtration $\{\mathcal{A}\}_{t\in\{1,\dots,H\}}$
- $c_t(\omega)\in\mathbb{R}^{n_t}$: cost coefficients
- $h_t(\omega)\in\mathbb{R}^{m_t}$: right-hand side parameters
- $W_t(\omega)\in\mathbb{R}^{m_t\times n_t}$: coefficients of $x_t(\omega)$
- $T_{t-1}(\omega)\in\mathbb{R}^{m_t\times n_{t-1}}$: coefficients of $x_{t-1}(\omega)$
- $x_t(\omega)$: set of state and action variables in period t

- We implicitly enforce non-anticipativity by requiring that x_t and ξ_t are adapted to filtration $\{\mathcal{A}\}_{t\in\{1,\dots,H\}}$
- $\forall A \in \mathcal{A}_k \setminus \mathcal{A}_{k-1}$, $x_t(\omega_1) = x_t(\omega_2)\ \forall \omega_1, \omega_2 \in A$

2.3.2 General formulation of the MSLP

The extended formulation of the MSLP is:

$$
\begin{aligned}
&\min c_1^T x_1 + \mathbb{E}[c_2(\omega)^T x_2(\omega) + \cdots + c_H(\omega)^T x_H(\omega)] \\
&s.t. \quad W_1 x_1 = h_1 \\
&T_1(\omega)x_1 + W_2(\omega)x_2(\omega) = h_2(\omega), \omega \in \Omega \\
&\qquad\qquad \vdots \\
&T_{t-1}(\omega)x_{t-1}(\omega) + W_t(\omega)x_t(\omega) = h_t(\omega), \omega \in \Omega \\
&\qquad\qquad \vdots \\
&T_{H-1}(\omega)x_{H-1}(\omega) + W_H(\omega)x_H(\omega) = h_H(\omega), \omega \in \Omega \\
&x_1 \geq 0, x_t(\omega) \geq 0, t = 2, \dots, H
\end{aligned}
\tag{2.3}
$$

We can now consider two specific instantiations of the MSLP: the scenario tree (MSLP-ST) and the lattice (MSLP-L). Using these notations:

- $\omega_t \in S_t$: index in the support Ξ_t of random input ξ_t
- $\omega_{[t]} \in S_1 \times \cdots \times S_t$ (interpretation: index in $\Xi_{[t]} = \Xi_1 \times \cdots \times \Xi_t$, which is the history of realizations, up to period t)

2.3.3 Scenario Tree formulation

$$
\begin{aligned}
&\min c_1^T x_1 + \mathbb{E}\left[c_2(\omega_{[2]})^T x_2(\omega_{[2]}) + \cdots + c_H(\omega_{[H]})^T x_H(\omega_{[H]})\right] \\
&s.t. \quad W_1 x_1 = h_1 \\
&T_1(\omega_{[2]})x_1 + W_2(\omega_{[2]})x_2(\omega_{[2]}) = h_2(\omega_{[2]}), \omega_{[2]} \in S_1 \times S_2 \\
&\qquad\qquad \vdots \\
&T_{t-1}(\omega_{[t]})x_{t-1}(\omega_{[t-1]}) + W_t(\omega_{[t]})x_t(\omega_{[t]}) = h_t(\omega_{[t]}), \omega_{[t]} \in S_1 \times \cdots \times S_t \\
&\qquad\qquad \vdots \\
&T_{H-1}(\omega_{[H]})x_{H-1}(\omega_{[H-1]}) + W_H(\omega_{[H]})x_H(\omega_{[H]}) = h_H(\omega_{[H]}), \omega_{[H]} \in S_1 \times \cdots \times S_H \\
&x_1 \geq 0, x_t(\omega_{[t]}) \geq 0, t = 2, \dots, H
\end{aligned}
\tag{2.4}
$$

2.3.4 Lattice formulation

$$
\begin{aligned}
&\min c_1^T x_1 + \mathbb{E}\left[c_2(\omega_2)^T x_2(\omega_{[2]}) + \cdots + c_H(\omega_H)^T x_H(\omega_{[H]})\right] \\
&s.t. \quad W_1 x_1 = h_1 \\
&T_1(\omega_2)x_1 + W_2(\omega_2)x_2(\omega_{[2]}) = h_2(\omega_2), \omega_{[2]} \in S_1 \times S_2 \\
&\qquad\qquad \vdots \\
&T_{t-1}(\omega_t)x_{t-1}(\omega_{[t-1]}) + W_t(\omega_t)x_t(\omega_{[t]}) = h_t(\omega_t), \omega_{[t]} \in S_1 \times \cdots \times S_t \\
&\qquad\qquad \vdots \\
&T_{H-1}(\omega_H)x_{H-1}(\omega_{[H-1]}) + W_H(\omega_H)x_H(\omega_{[H]}) = h_H(\omega_H), \omega_{[H]} \in S_1 \times \cdots \times S_H \\
&x_1 \geq 0, x_t(\omega_{[t]}) \geq 0, t = 2, \ldots, H
\end{aligned}
\tag{2.5}
$$

→ Note: There exists some relations to other decision making problems such as statistical decision theory, dynamic programming, online optimization and stochastic control.

Performance

3.1 Notations

Using (2.1), let's define the following:

- $z(x, \xi) = c^T x + Q(x, \xi) + \delta(x|K_1)$
- $Q(x, \xi) = \min_y \{q(\omega)^T y \mid W(\omega)y = h(\omega) - T(\omega)x\}$
- $K_1 = \{x | Ax = b, x \geq 0\}$ is the set of feasible first-stage decisions
- $K_2(\omega) = \{x \mid \exists y \geq 0 : W(\omega)y = h(\omega) - T(\omega)x\}$ is the set of first-stage decisions that have a feasible reaction in the second stage for $\omega \in \Omega$
- It is possible that $z(x, \xi) = +\infty$ (if $x \notin K_1 \cap K_2(\omega)$)
- It is possible that $z(x, \xi) = -\infty$ (unbounded below)

3.2 Expected value of perfect information

There are 2 tactics:

- **wait-and-see** value is the expected value of reacting with perfect foresight (we know everything that will happen) $x^*(\xi)$ to ξ:

$$WS = \mathbb{E}[\min_x z(x, \xi)] = \mathbb{E}[z(x^*(\xi), \xi)] \tag{3.1}$$

- **here-and-now** value is the expected value of the recourse problem (remove non-anticipativity constraint):

$$SP = \min_x \mathbb{E}[z(x, \xi)] \tag{3.2}$$

The **expected value of perfect information** is like the value we give to getting a perfect forecast for the future and is thus defined like this:

$$EVPI = SP - WS \tag{3.3}$$

3.3 The value of the stochastic solution

Here too there are 2 tactics:

- **expected value problem**
$$EV = \min_x z(x, \bar{\xi}) = \mathbb{E}[\xi] \tag{3.4}$$
and its **expected value solution** is noted $x^*(\bar{\xi})$.
- **expected value of using the EV solution** measures the performance of $x^*(\bar{\xi})$:
$$EEV = \mathbb{E}[z(x^*(\bar{\xi}), \xi)] \tag{3.5}$$

The **value of the stochastic solution** is noted like this:
$$VSS = EEV - SP \tag{3.6}$$

3.4 Basic inequalities

3.4.1 Crystal Ball

For every ξ, we have $z(x^*(\xi), \xi) \leq z(x^*, \xi)$ where x^* is the optimal solution to the stochastic program. And if we take the expectation of this inequality, we have $WS \leq SP$, because WS is a relaxation. It explains that we can do better with a crystal ball.

3.4.2 Lazy solution

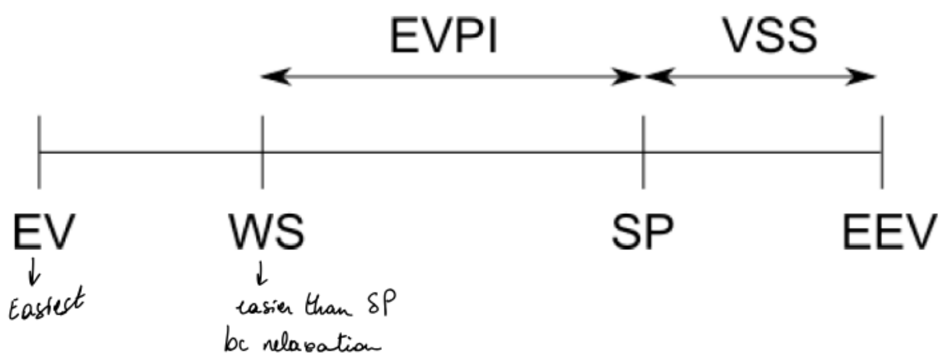
Knowing that x^* is the optimal solution of $\min_x \mathbb{E}[z(x, \xi)]$ and $x^*(\bar{\xi})$ is a solution but not necessarily optimal then we have $SP \leq EEV$, because:
$$\min_x \mathbb{E}[z(x, \xi)] = SP \leq EEV = \mathbb{E}[z(x^*(\bar{\xi}), \xi)] \tag{3.7}$$

3.4.3 Link between all the values

We know that:

- $VSS \geq 0$
- $EVPI \geq 0$
- $VSS \leq EEV - EV$
- $EVPI \leq EEV - EV$
- If $EEV - EV = 0$ then $VSS = EVPI = 0$

and the inequalities can be summarized in the following diagram:



3.5 Bounds on EVPI and VSS

First let's introduce the pairs subproblem of ξ^r and ξ^k:

$$\begin{aligned} \min z^P(x, \xi^r, \xi^k) &= c^T x + p^r q^T y(\xi^r) + (1 - p^r) q^T y(\xi^k) \\ s.t.\ Ax &= b \\ Wy(\xi^r) &= \xi^r - Tx \\ Wy(\xi^k) &= \xi^k - Tx \\ x, y &\geq 0 \end{aligned} \tag{3.8}$$

- $(\bar{x}^k, \bar{y}^k, y(\xi^r))$ denotes an optimal solution to the problem and z^k is the optimal objective function value $z^P(\bar{x}^k, \bar{y}^k, y(\xi^k))$
- $z^P(x, \xi^r, \xi^r)$ corresponds to the deterministic optimization against the reference scenario
- if $\xi^r \notin \Xi$, $p^r = 0$ and $z^P(x, \xi^r, \xi^k) = z(x, \xi^k)$

The **sum of pairs expected value (SPEV)**:

$$SPEV = \frac{1}{1 - p^r} \sum_{k=1, k \neq r}^{K} p^k \min z^P(x, \xi^r, \xi^k) \tag{3.9}$$

When $\xi^r \notin \Xi$ then $SPEV = WS$: When $p^r = 0, z^P(x, \xi^r, \xi^k)$ coincides with $z(x, \xi^k)$. Therefore $SPEV = \sum_{k=1}^{K} p^k \min_x z(x, \xi^k) = WS$.
We then know $WS \leq SPEV \leq SP$.

3.5.1 Upper bound on SP: EVRS and EPEV

- The **expected value of the reference scenario** is $EVRS = \mathbb{E}_{\xi}(\bar{x}^r, \xi)$, where $\bar{x}^r$ is the optimal solution to $z(x, \xi^r)$.
- The **expectation of pairs of expected value** is defined as

  $$EPEV = \min_{k=1,\ldots,K \cup \{r\}} \mathbb{E}_{\xi}(\bar{x}^r, \xi)$$

  where $(\bar{x}^k, \bar{y}^k, y(\xi^K))$ is the optimal solution to the pairs subproblem of ξ^r and ξ^k.

As $SP, EPEV, EVRS$ are the optimal values of $\min_x \mathbb{E}_{\xi} z(x, \xi)$ over smaller feasible sets:

$$SP \leq EPEV \leq EVRS \tag{3.10}$$

Because

- SP: $x \in K_1 \cap K_2$
- $EPEV$: $x \in K_1 \cap K_2 \cap \{\bar{x}^k, k = 1, \ldots, K \cup \{r\}\}$
- $EVRS$: $x \in \bar{x}^r \cap K_1 \cap K_2$

3.6 Estimations of WS and EEV

An estimation of WS and EEv can be done through a sample mean approximation: from samples $\xi_i = \xi(\omega_i)$ for $i = 1, \ldots, K$,

1. Compute $x^\star(\bar{\xi})$;
2. Compute $WS_i = z(x^\star(\xi_i), \xi_i)$ and $EEV_i = c^T x^\star(\xi_i) + Q(x^\star(\bar{\xi}), \xi_i)$;
3. Estimate $\bar{WS} = \frac{1}{K}\sum_{i=1}^{K} WS_i$ and $E\bar{E}V = \frac{1}{K}\sum_{i=1}^{K} EEV_i$.

3.6.1 Central Limit Theorem

Suppose $\{X_1, \ldots, X_K\}$ is a sequence of iid rv with $\mathbb{E}[X_i] = \mu$ and $Var[X_i] = \sigma^2 < \infty$. Then, as n approaches infinity, $\sqrt{n}(S_n - \mu)$ converge in distribution to a normal $\mathcal{N}(0, \sigma^2)$:

$$\sqrt{n}\left(\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - \mu\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2) \tag{3.11}$$

The central limit theorem is useful to decrease the importance of rare but extreme events.

3.6.2 Importance sampling

Suppose we wish to estimate $\mathbb{E}[C(\omega)]$, where ω is distributed according to $f(\omega)$ and estimates $\mathbb{E}[C(\omega)]$ with $\sum_{i=1}^{N} \frac{1}{N} C(\omega_i)$. A sample average pulls samples ω_i according to the distribution function $f(\omega)$, while the importance sampling pulls the samples ω_i according to the distribution $g(\omega) = \frac{f(\omega)C(\omega)}{\mathbb{E}[C(\omega)]}$, where the $\mathbb{E}[C(\omega)]$ is an approximation of the real expectation. It then estimates $\mathbb{E}[C(\omega)]$ with $\sum_{i=1}^{N} \frac{1}{N} \frac{f(\omega_i)C(\omega_i)}{g(\omega_i)}$.

Benders Decomposition

4.1 Cutting plane methods

A cutting plane method is an optimisation method based on the idea of iteratively refining the objective function, or a set of feasible constraints of a problem through linear inequalities (see LINMA2450).

4.1.1 Nomenclature

- The benders decomposition is a specific method for obtaining the cutting planes when $F(x)$ is the value function of a second-stage linear program.
- The L-shaped method is a specific instance of Benders decomposition when the second-stage linear program is decomposable into a set of scenarios.
- The multi-cut L-shaped method is an alternative to the L-shaped method which generates multiple cutting planes at step 1 of Kelley's method (see 4.1.2).

4.1.2 Kelley's Cutting Plane Algorithm

This algorithm is designed to solve convex but non-differentiable optimization problems of the form

$$\begin{aligned} z^\star = \min c^T x + F(x) \\ \text{s.t. } x \in X \end{aligned} \tag{4.1}$$

where $X \subseteq \mathbb{R}^n$ is convex and compact, $F : \mathbb{R}^n \to \mathbb{R}$ is convex and $c \in \mathbb{R}^n$ is a parameter vector.
Let us define

- $L_k : \mathbb{R}^n \to \mathbb{R}$ a lower bound function of $F(x)$ at iteration k;
- A lower bound L_k of $z^\star$ at iteration k;
- An upper bound U_k of $z^\star$ at iteration k.

Algorithm 1 Kelley's Cutting plane algorithm

1: **Step 0:** Set $k = 0$ and assume $x_1 \in X$ is given. Set $L_0(x) = -\infty$ for all $x \in X$, $U_0 = c^T x_1 + F(x_1)$, and $L_0 = -\infty$.

2: **Step 1:** Set $k = k + 1$. Find $a_k \in \mathbb{R}$ and $b_k \in \mathbb{R}^n$ such that

$$F(x_k) = a_k + b_k^T x_k$$

$$F(x) \geq a_k + b_k^T x \qquad x \in X$$

3: **Step 2:** Set

$$U_k = \min(U_{k-1}, c^T x_k + F(x_k))$$

and

$$L_k(x) = \max(L_{k-1}(x), a_k + b_k^T x) \qquad x \in X$$

4: **Step 3:** Compute

$$L_k = \min_{x \in X} L_k(x) + c^T x$$

and denote x_{k+1} as the optimal solution of this problem.

5: **Step 4:** If $U_k - L_k = 0$, stop. Otherwise, go to step 1.

4.2 Context and description

Consider the following optimization problem:

$$\begin{aligned} z^\star &= \min c^T x + q^T y \\ Ax &= b \\ Tx + Wy &= h \\ x, y &\geq 0 \end{aligned} \tag{4.2}$$

with $x \in \mathbb{R}^{n_1}$, $y \in \mathbb{R}^{n_2}$, $c \in \mathbb{R}^{n_1}$, $q \in \mathbb{R}^{n_2}$, $A \in \mathbb{R}^{m_1 \times n_1}$, $b \in \mathbb{R}^{m_1}$, $T \in \mathbb{R}^{m_2 \times n_1}$, $W \in \mathbb{R}^{m_2 \times n_2}$, $h \in \mathbb{R}^{m_2}$[1].

We use Benders decomposition when the entire problem is difficult to solve, and if the constraint $Tx + Wy = h$ is ignored, the problem becomes easy to solve, or if fixing x simplifies the computation of the solution.

4.2.1 Idea of Benders decomposition

Define the value function $V : \mathbb{R}^{n_1} \to \mathbb{R}$:

$$\begin{aligned} V(x) &= \min_y \{q^T y\} \\ Wy &= h - Tx \\ y &\geq 0 \end{aligned} \tag{4.3}$$

[1] It is not necessarily a stochastic problem

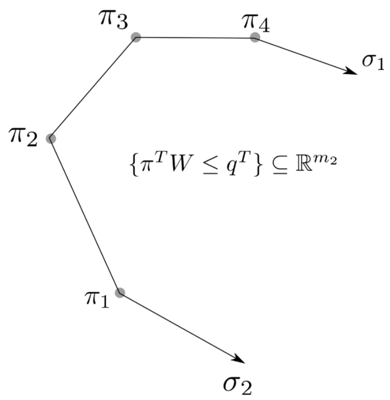
Or equivalently,

$$\begin{aligned} \min c^T x + V(x) \\ Ax = b \\ x \in dom(V) \\ x \geq 0 \end{aligned} \tag{4.4}$$

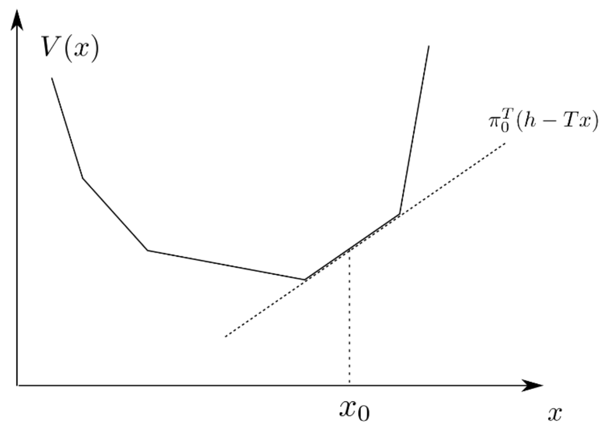
where $dom(V) = \{x \in \mathbb{R}^{n_1} \mid \exists y \geq 0 : Wy = h - Tx\}$.
The dual of (4.3) is

$$\begin{aligned} \max_{\pi} \pi^T(h - Tx) \\ \pi^T W \leq q^T \end{aligned} \tag{4.5}$$

Let us call E the set of extreme points of $\pi^T W \leq q^T$ and R the set of extreme rays of $\pi^T W \leq q^T$ (see (1.8) for definitions).



We can see that $V(x)$ is a piecewise linear convex function of x and, defining x_0 as the dual optimal multiplier of (4.3) given x_0, then $\pi_0^T(h - Tx_0)$ is a supporting hyperplane of $V(x)$ at x_0, because it belongs to the subdifferential of $V(x)$ at $h - Tx_0$.



From this, we can also express the domain of V as follows:

$$dom(V) = \{x \mid \sigma^T(h - Tx) \leq 0,\ \sigma \in R\} \tag{4.6}$$

where $\sigma \in R$ is the set of extreme rays of $\pi^T W \leq q^T$.