L-shaped Method from the Dual View

Zhuobie

2026 年 4 月 10 日

1. The Main Idea of the L-shaped Method

We consider the two-stage stochastic linear program with finitely scenarios:

$$\min_x \ c^T x + \sum_{k=1}^{K} p_k Q_k(x)$$

subject to

$$Ax \le b, \qquad x \ge 0,$$

where, for each scenario k,

$$Q_k(x) = \min_{y_k} \ q_k^T y_k$$

subject to

$$W y_k = h_k - T_k x, \qquad y_k \ge 0.$$

The purpose of the L-shaped method is to avoid solving the full extensive form directly. Instead, we exploit the block structure of the second-stage problems and build a master problem only in the first-stage variable x together with cuts generated from the scenario subproblems. This is the standard decomposition idea behind the L-shaped method.

Guiding idea. The extensive form contains one copy of the recourse system for each scenario. The L-shaped method projects the problem onto the first-stage space and recovers the second-stage effect through cuts.

Next, we project the original problem into a more compact form.

If for each x we solve the scenario subproblem and denote an optimal solution by $\bar{y}_k(x)$, then the original problem can be viewed in compact form as

$$\min_x \ c^T x + \sum_{k=1}^{K} p_k q_k^T \bar{y}_k(x)$$

subject to

$$Ax \le b, \qquad x \ge 0.$$

Equivalently, the only nonlinear-looking part is the recourse value function $Q_k(x)$. Thus the entire method reduces to understanding how to describe $Q_k(x)$ by linear inequalities in the first-stage space.

Remark

This projection viewpoint is exactly why the L-shaped method feels natural in stochastic programming: the first-stage variable x is the linking variable, while each scenario contributes one recourse function $Q_k(x)$.

2. Dual of the Scenario Subproblem

For a fixed scenario k, the primal recourse problem is

$$(P_k(x)) \qquad \min_{y_k} \; q_k^T y_k$$

subject to

$$W y_k = h_k - T_k x, \qquad y_k \geq 0.$$

Its dual is

$$(DP_k(x)) \qquad \max_{\lambda_k} \; (h_k - T_k x)^T \lambda_k$$

subject to

$$W^T \lambda_k \leq q_k.$$

If $(DP_k(x))$ has a finite optimal solution $\bar{\lambda}_k(x)$, then by strong duality,

$$Q_k(x) = q_k^T \bar{y}_k(x) = (h_k - T_k x)^T \bar{\lambda}_k(x).$$

The crucial observation is that the feasible region of the dual subproblem,

$$\mathcal{D}_k := \{\lambda_k \mid W^T \lambda_k \leq q_k\},$$

depends only on the problem data and *does not depend on* x. This is the key structural step: after dualization, the dependence on x appears only in the objective, not in the constraints.

Why is this so important? Once the feasible set of the dual is fixed, we can study its extreme points and extreme rays once and for all. Then changing x only changes which extreme point becomes optimal, or whether some extreme ray makes the dual unbounded.

3. Decomposition the dual variables by Extreme Points and Extreme Rays

Since $\mathcal{D}_k = \{\lambda_k \mid W^T \lambda_k \leq q_k\}$ is a polyhedron, every feasible λ_k can be represented as a convex combination of extreme points plus a conic combination of extreme rays.

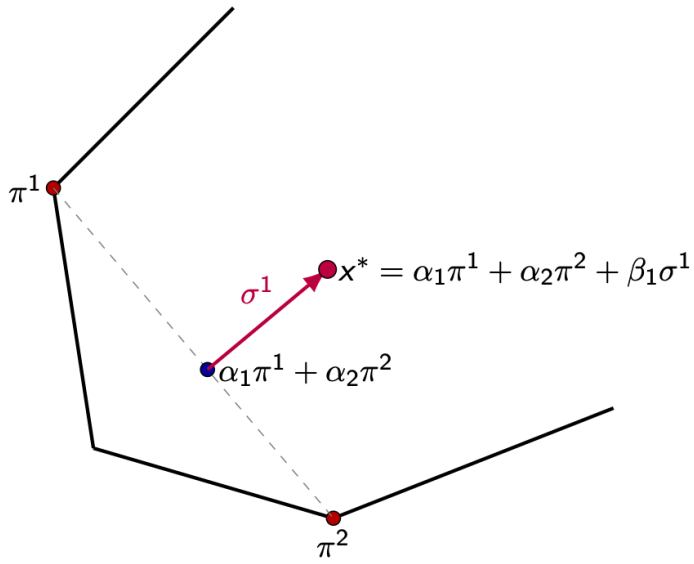
Let

$$\{\pi_k^\ell : \ell = 1, \ldots, m_k^1\}$$

be the extreme points of $\mathcal{D}_k$, and let

$$\{\sigma_k^\ell : \ell = 1, \ldots, m_k^2\}$$

be the extreme rays of $\mathcal{D}_k$.



Then any feasible λ_k can be written as

$$\lambda_k = \sum_{\ell=1}^{m_k^1} \alpha_\ell \pi_k^\ell + \sum_{\ell=1}^{m_k^2} \beta_\ell \sigma_k^\ell,$$

with

$$\sum_{\ell=1}^{m_k^1} \alpha_\ell = 1, \qquad \alpha_\ell \geq 0, \qquad \beta_\ell \geq 0.$$

Substituting this into the dual objective yields

$$\max \sum_{\ell=1}^{m_k^1} \alpha_\ell (h_k - T_k x)^T \pi_k^\ell + \sum_{\ell=1}^{m_k^2} \beta_\ell (h_k - T_k x)^T \sigma_k^\ell$$

subject to

$$\sum_{\ell=1}^{m_k^1} \alpha_\ell = 1, \qquad \alpha_\ell \geq 0, \qquad \beta_\ell \geq 0.$$

4. Two Possible Situations in the Dual Subproblem

From the representation above, two cases may occur.

4.1 Case A: The Dual is Unbounded —> feasible cuts

Suppose there exists an extreme ray σ_k^ℓ such that

$$(h_k - T_k x)^T \sigma_k^\ell > 0.$$

Then by increasing the corresponding coefficient β_ℓ, the dual objective goes to $+\infty$. Hence $(DP_k(x))$ is unbounded.
By duality, this means the primal subproblem $(P_k(x))$ is infeasible.

Therefore, if we want the primal subproblem to be feasible, we must impose

$$(h_k - T_k x)^T \sigma_k^\ell \leq 0 \qquad \forall \ell = 1, \ldots, m_k^2.$$

This can be rewritten as

$$(\sigma_k^\ell)^T T_k x \geq (\sigma_k^\ell)^T h_k.$$

These are the **feasibility cuts**. They exclude first-stage decisions x for which the kth recourse problem is infeasible.

Remark

An extreme ray does not give an optimal value. Instead, it detects infeasibility of the primal recourse problem. This is why rays generate feasibility cuts rather than value approximations.

4.1.1 Violation Reformulation for Feasibility Cuts

Recall that at iteration ν, for scenario k, feasibility is checked through the auxiliary problem

$$\min \; w' = e^T u^+ + e^T u^-$$

subject to

$$Wy + Iu^+ - Iu^- = h_k - T_k x^\nu,$$
$$y \geq 0, \qquad u^+ \geq 0, \qquad u^- \geq 0,$$

where $e = (1, \ldots, 1)^T$.

If the optimal value satisfies $w' > 0$, then the recourse subproblem is infeasible, and a feasibility cut must be generated. The dual of the auxiliary problem is

$$\max_{\sigma_k} (h_k - T_k x^\nu)^T \sigma_k$$

subject to

$$W^T \sigma_k \leq 0, \qquad -e \leq \sigma_k \leq e.$$

If the optimal value of $(V_k(x^\nu))$ is positive, then there exists a direction σ_k that violates recourse feasibility, and therefore the primal recourse problem is infeasible.

The box constraints

$$-e \leq \sigma_k \leq e$$

play the role of a *normalization*. They cut the recession cone by a bounded set, so instead of returning an arbitrarily scaled ray, the LP returns a normalized representative of that ray.

4.2 Case B: The Dual is Bounded and Has a Finite Optimum —> optimality cuts

Now suppose

$$(h_k - T_k x)^T \sigma_k^\ell \leq 0 \qquad \forall \ell = 1, \ldots, m_k^2.$$

Then no extreme ray can improve the objective indefinitely, so the dual is bounded above.

In that case, the optimum is attained at an extreme point of $\mathcal{D}_k$. Hence there exists some ℓ^* such that

$$Q_k(x) = (h_k - T_k x)^T \pi_k^{\ell^*}.$$

To describe the optimal value of the recourse problem inside the master problem, we introduce a variable θ_k intended to represent $Q_k(x)$).

Since for any extreme point π_k^ℓ we must have

$$\theta_k \geq (h_k - T_k x)^T \pi_k^\ell, \qquad \ell = 1, \ldots, m_k^1,$$

we obtain the **optimality cuts**.

Equivalently,

$$\theta_k + (\pi_k^\ell)^T T_k x \geq (\pi_k^\ell)^T h_k, \qquad \ell = 1, \ldots, m_k^1.$$

Why the Optimal Extreme Point Changes When x Changes A useful geometric interpretation is the following.

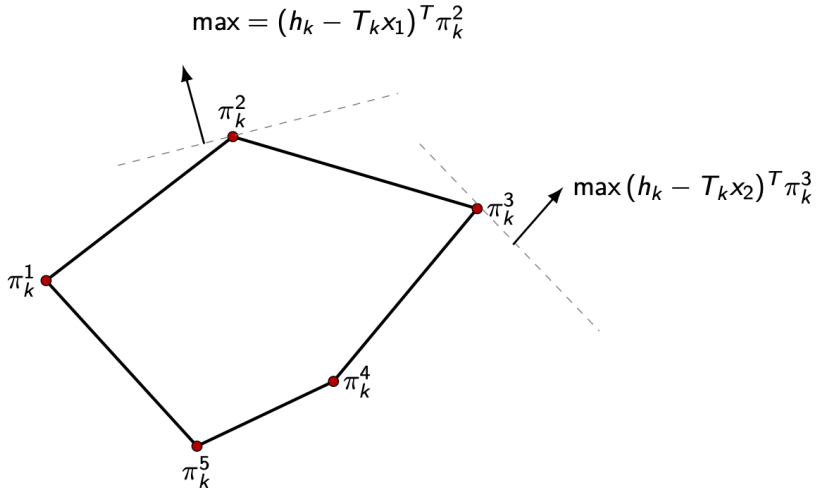
The dual feasible region $\mathcal{D}_k$ is fixed. As x changes, only the objective vector

$$h_k - T_k x$$

changes. Therefore, the maximizing point in

$$\max_{\lambda_k \in \mathcal{D}_k} (h_k - T_k x)^T \lambda_k$$

moves from one extreme point to another.



So the recourse value function is not linear globally, but piecewise linear: different regions of the first-stage space correspond to different active extreme points.

This is exactly why a cutting-plane method works: every time we solve the subproblem at a current iterate, we reveal one more supporting hyperplane of the true recourse function.

5. Master Problem Induced by the Cuts

Putting together the first-stage constraints, the feasibility cuts, and the optimality cuts, we obtain the master problem

$$\min \; c^T x + \sum_{k=1}^{K} p_k \theta_k$$

subject to

$$Ax = b, \qquad x \geq 0,$$

together with

Feasibility cuts

$$(\sigma_k^\ell)^T T_k x \geq (\sigma_k^\ell)^T h_k, \qquad k = 1, \ldots, K, \quad \ell = 1, \ldots, m_k^2,$$

and

Optimality cuts

$$\theta_k \geq (h_k - T_k x)^T \pi_k^\ell, \qquad k = 1, \ldots, K, \quad \ell = 1, \ldots, m_k^1.$$

In principle, this description is exact. In practice, however, one does not enumerate all extreme points and extreme rays in advance. Instead, the L-shaped method generates only those cuts that are violated at the current trial point.

6. Single-Cut Interpretation

In many presentations of the L-shaped method, one aggregates the scenario optimality information into a single variable

$$\Theta \approx \sum_{k=1}^{K} p_k Q_k(x).$$

If for the current iterate x^ν we obtain dual optimal extreme points π_k^ν for each scenario, then we can write the aggregated optimality cut as

$$\Theta + \sum_{k=1}^{K} p_k (\pi_k^\nu)^T T_k x \geq \sum_{k=1}^{K} p_k (\pi_k^\nu)^T h_k.$$

This is the classical *single-cut* version described in the lecture notes: the expected recourse function is approximated by one supporting hyperplane per iteration.

Remark

The multi-cut version keeps separate variables θ_k and adds one optimality cut per scenario. The single-cut version aggregates them into one variable Θ and one cut per iteration.

7. L-shaped Method Code & Examples

This section records two standard examples for feasibility cuts and optimality cuts, together with the implementation reference used in practice.

7.1 Feasibility Cut

We first consider the following two-stage stochastic program:

$$\min\ 3x_1 + 2x_2 - \mathbb{E}_\xi(15y_1 + 12y_2)$$

subject to

$$\begin{aligned} 3y_1 + 2y_2 &\le x_1, \\ 2y_1 + 5y_2 &\le x_2, \\ 0.8\xi_1 \le y_1 &\le \xi_1, \\ 0.8\xi_2 \le y_2 &\le \xi_2, \\ x, y \ge 0 \qquad &\text{a.s.} \end{aligned}$$

where

$$\xi_1 = \begin{cases} 4, & p = 0.5, \\ 6, & p = 0.5, \end{cases} \qquad \xi_2 = \begin{cases} 4, & p = 0.5, \\ 8, & p = 0.5. \end{cases}$$

This example is a standard illustration of *feasibility cut generation*: when a trial first-stage decision makes the recourse problem infeasible for some scenario, an auxiliary feasibility problem is solved and a cut is generated to eliminate the current infeasible point.

Remark

Reference example: Example 4.2, p. 192, *Introduction to Stochastic Programming*.

7.2 Optimality Cut

We next consider the following problem for illustrating optimality cuts:

$$z = \min\ 100x_1 + 150x_2 + \mathbb{E}_\xi(q_1y_1 + q_2y_2)$$

subject to

$$\begin{aligned} x_1 + x_2 &\le 120, \\ 6y_1 + 10y_2 &\le 60x_1, \\ 8y_1 + 5y_2 &\le 80x_2, \\ y_1 \le d_1, \qquad y_2 &\le d_2, \\ x_1 \ge 40, \qquad x_2 \ge 20, \qquad y_1, y_2 &\ge 0. \end{aligned}$$

The random vector is

$$\xi^T = (d_1, d_2, q_1, q_2),$$

with

$$\begin{aligned} p = 0.4: &\quad (500, 100, -24, -28), \\ p = 0.6: &\quad (300, 300, -28, -32). \end{aligned}$$

This example shows the role of *optimality cuts*: once the recourse problem is feasible, the dual optimal multipliers

define a supporting hyperplane of the expected recourse function, which is then added to the master problem.

Remark

Reference example: Example 1, p. 184, *Introduction to Stochastic Programming*.

References

- Birge, J. R., and Louveaux, F. *Introduction to Stochastic Programming*.
- GitHub repository for L-shaped method code and examples: `https://github.com/Zhuobie2025/Multi_stage_stochastic_programming`