# Cutting Plane Methods

## Operations Research

Anthony Papavasiliou

# Contents

# Table of Contents

# Cutting Plane Methods

**Cutting plane methods**: optimization methods which are based on the idea of iteratively refining the objective function or set of feasible constraints of a problem through linear inequalities

# Kelley's Cutting Plane Algorithm

Kelley's cutting plane algorithm is designed for solving convex non-differentiable optimization problems:

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

where

- $X$ is a compact convex subset of $\mathbb{R}^n$
- $F : \mathbb{R}^n \to \mathbb{R}$ is a convex function
- $c \in \mathbb{R}^n$ is a parameter vector

# Kelley's Cutting Plane Algorithm

Define

- $L_k : \mathbb{R}^n \to \mathbb{R}$ as lower bounding function of $F(x)$ at iteration $k$
- Lower bound $L_k$ of $z^\star$ at iteration $k$
- Upper bound $U_k$ of $z^\star$ at iteration $k$

Idea: gradually bound $F(x)$ from below with functions $L_k(x)$

# Kelley's Cutting Plane Algorithm

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

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

$$F(x_k) = a_k + b_k^T x_k$$
$$F(x_k) \geq a_k + b_k^T x, x \in X$$

*Step 2*: Set

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

and

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

# Kelley's Cutting Plane Algorithm

*Step 3*: Compute

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

and denote $x_k$ as the optimal solution of this problem

*Step 4*: If $U_k - L_k = 0$, stop; else, repeat from step 1

# Nomenclature of Cutting Plane Methods

- **Benders decomposition**: specific method for obtaining the cutting planes when $F(x)$ is the value function of a second-stage linear program
- **L-shaped method**: specific instance of Benders decomposition when second-stage linear program is decomposable into a set of scenarios
- **Multi-cut L-shaped method**: alternative to L-shaped method which generates multiple cutting planes at step 1 of Kelley's method
- Cutting plane methods generalized to **bundle methods** in non-differentiable convex optimization (commonly used in Lagrange relaxation)

# Table of Contents

# When to Use Benders Decomposition

Consider the following optimization problem:

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

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

- This is not (necessarily) a stochastic program
- This is a two-stage program

Context for Benders decomposition:

1. entire problem is difficult to solve
2. if $Tx + Wy = h$ is ignored, problem is relatively easy
3. if $x$ is fixed, problem is relatively easy

# Idea of Benders Decomposition

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

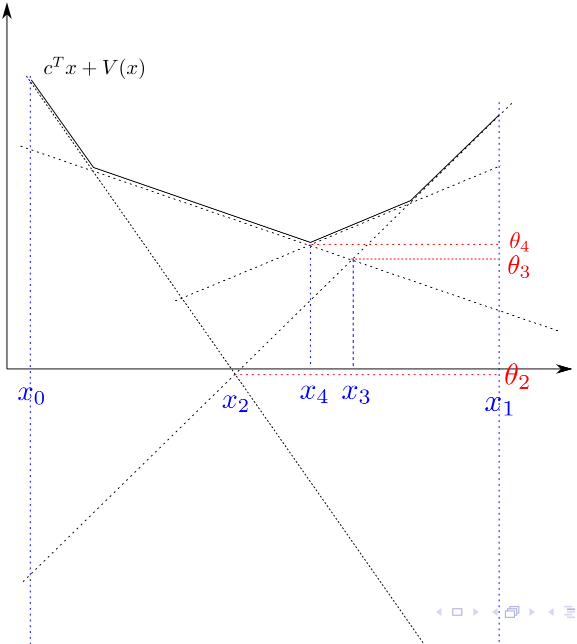
$$
\begin{aligned}
(S): \quad & V(x) = \min_y q^T y \\
& Wy = h - Tx \\
& y \geq 0
\end{aligned}
$$

Equivalent description of problem

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

Note: $\operatorname{dom} V = \{x : \exists y, Tx + Wy = h, y \geq 0\}$

# Graphical Description of Benders Decomposition



# Table of Contents

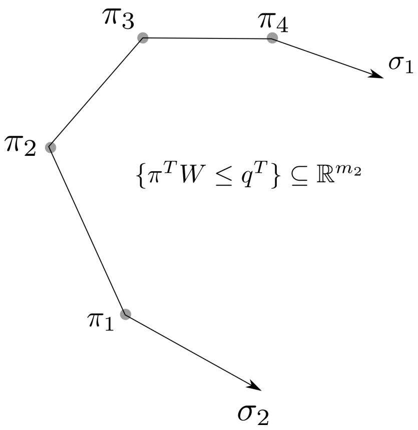
# Dual of Second-Stage Linear Program

The dual of ($S$) can be expressed as:

$$(D): \quad \max_{\pi} \pi^T (h - Tx)$$
$$\pi^T W \leq q^T$$

Note: feasible region of ($D$) does *not* depend on $x$

- $V$: set of extreme points of $\pi^T W \leq q^T$
- $R$: set of extreme rays of $\pi^T W \leq q^T$




$\pi \in V, \sigma \in R$ do *not* depend on $x$, can be enumerated

# Value Function Is Piecewise Linear

- $V(x)$ is a piecewise linear convex function of $x$
- If $\pi_0$ is dual optimal multiplier of $(S)$ given $x_0$, then

  $$\pi_0^T(h - Tx_0)$$

  is a supporting hyperplane of $V(x)$ at $x_0$

We recall a previous result for the proof

# Parametrizing the Right-Hand Side

Define $c(u)$ as optimal value of

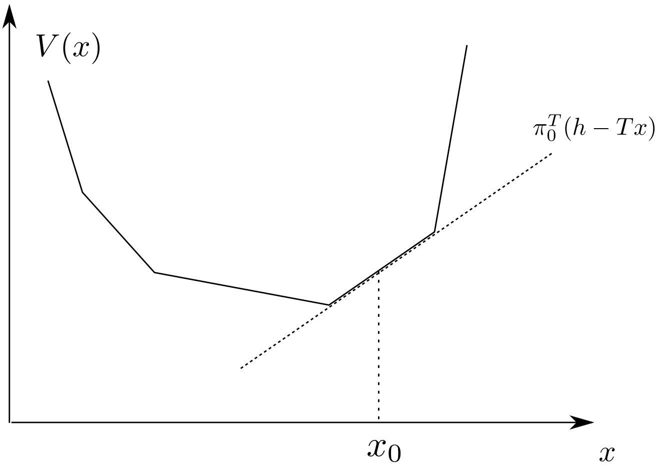
$$c(u) = \min f_0(x)$$
$$f_i(x) \leq u_i, i = 1, \ldots, m$$

where $x \in \text{dom } f_0$ is the convex domain of $f_0(x)$ and $f_0, f_i$ are convex functions

- $c(u)$ is convex
- Suppose strong duality holds and denote $\lambda^\star$ as the maximizer of the dual function $\inf_{x \in \text{dom } f_0}(f_0(x) - \lambda^T(f(x) - u))$ for $\lambda \leq 0$. Then $\lambda^\star \in \partial c(u)$.

From previous result:

- $V(h - Tx)$ is convex, so $V(x)$ is convex
- $\pi_0 \in \partial V(h - Tx_0)$, so $\pi_0^T(h - Tx)$ is a supporting hyperplane of $V(x)$ at $x_0$
- $(S)$ has a finite number of dual optimal multipliers $\Rightarrow$ finite number of supporting hyperplanes for $V(x) \Rightarrow V(x)$ is piecewise linear convex


$V(x)$
$\pi_0^T(h - Tx)$
$x_0$
$x$

# Domain of Value Function

dom $V$ can be expressed equivalently as follows:

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

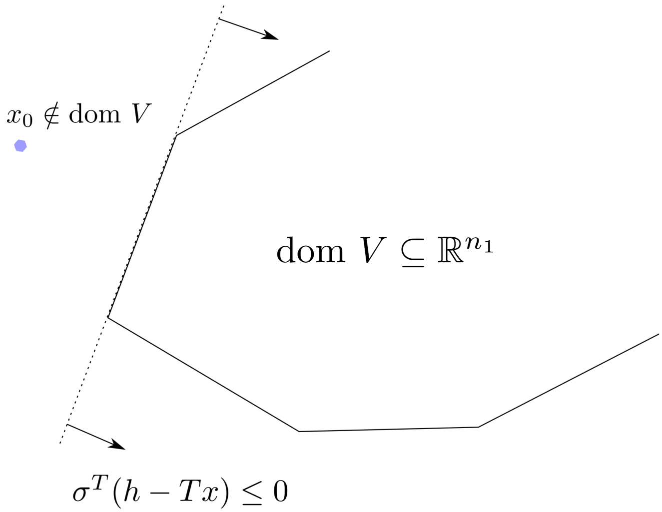
where $\sigma \in R$ is the set of extreme rays of $\pi^T W \leq q^T$

Proof that dom $V \subseteq \{\sigma^T(h - Tx) \leq 0, \sigma \in R\}$:

- Suppose $x \in$ dom $V$ and $\sigma^T(h - Tx) > 0$ for some $\sigma \in R$
- $\sigma$ is an extreme ray $\Rightarrow \sigma^T W \leq 0$
- Consider any dual feasible vector $\pi_0$: $\pi_0 + \lambda\sigma$ is feasible for any $\lambda \geq 0$
- Since $\sigma^T(h - Tx) > 0$, $(D)$ becomes unbounded
- Contradiction with assumption that $x \in$ dom $V \Rightarrow$ $\sigma^T(h - Tx) \leq 0$ for all $\sigma \in R$

Proof that $\{\sigma^T(h - Tx) \leq 0, \sigma \in R\} \subseteq \text{dom } V$:

- Any ray of $\pi^T W \leq q^T$ can be expressed as convex combination of extreme rays
- Therefore, for any ray $\sigma$ of $\pi^T W \leq q^T$ it follows that $\sigma^T(h - Tx) \leq 0 \Rightarrow (D)$ cannot become unbounded


$x_0 \notin \text{dom } V$
$\text{dom } V \subseteq \mathbb{R}^{n_1}$
$\sigma^T(h - Tx) \leq 0$

# Reformulation

$$\min c^T x + \theta$$
$$Ax = b$$
$$\sigma_r^T (h - Tx) \leq 0, \sigma_r \in R$$
$$\theta \geq \pi_v^T (h - Tx), \pi_v \in V$$
$$x \geq 0$$

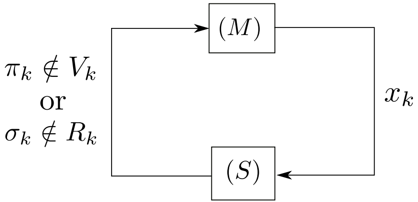
$\theta$: free auxiliary variable

# Master Problem

Relax inequalities that define $V(x)$ and dom $V$:

$$
\begin{aligned}
(M): \quad & z_k = \min c^T x + \theta \\
& Ax = b \\
& \sigma^T(h - Tx) \leq 0, \sigma \in R_k \subseteq R \\
& \theta \geq \pi^T(h - Tx), \pi \in V_k \subseteq V \\
& x \geq 0
\end{aligned}
$$

# Bounds and Exchange of Information




Solution of master problem provides:

- lower bound $z_k \leq z^\star$
- candidate solution $x_k$
- under-estimator of $V(x_k)$, $\theta_k \leq V(x_k)$

Solution of slave problem with input $x_k$ provides:

- upper bound $c^T x_k + q^T y_{k+1} \geq z^\star$
- new vertex $\pi_{k+1}$ or new extreme ray $\sigma_{k+1}$

# Table of Contents

# Benders Decomposition Algorithm

*Step 0*: Set $k = 0$, $V_0 = R_0 = \emptyset$.

*Step 1*: Solve $(M)$. Store $x_k$.

- If $(M)$ is feasible, store $x_k$.
- If $(M)$ is infeasible, exit. Problem is infeasible.

*Step 2*: Solve $(S)$ with $x_k$ as input.

- If $(S)$ is infeasible, let $R_{k+1} = R_k \cup \{\sigma_{k+1}\}$. Let $k = k + 1$ and return to step 1.
- If $(S)$ is feasible, let $V_{k+1} = V_k \cup \{\pi_{k+1}\}$
  - If $V_k = V_{k+1}$, terminate with $(x_k, y_{k+1})$ as optimal solution.
  - Else, let $k = k + 1$ and return to step 1.

*Finite termination* since $V$ and $R$ are finite

# Proof of Convergence

Denote $x_k$ as solution of $(M)$ and use it as input in $(S)$

- Suppose $(S)$ is feasible, denote $\pi_{k+1}$ as optimal vertex. If $\pi_{k+1} \in V_k$ then $x_k$ is optimal.
- Suppose $(S)$ is infeasible, denote $\sigma_{k+1}$ as extreme ray. Then $\sigma_{k+1} \notin R_k$.

Proof that $\pi_{k+1} \in V_k \Rightarrow x_k$ is optimal

- For any $x$ feasible, $c^T x + V(x) \geq c^T x_k + \theta_k$ because $(M)$ is a relaxation of the original problem
- If $\theta_k = V(x_k)$, then $x_k$ is optimal since for any $x$ feasible, $c^T x + V(x) \geq c^T x_k + V(x_k)$
- We already know that $\theta_k \leq V(x_k)$ (first bullet)
- Need to show that $\theta_k \geq V(x_k)$ (next slide)

Proof that $\pi_{k+1} \in V_k \Rightarrow \theta_k \geq V(x_k)$

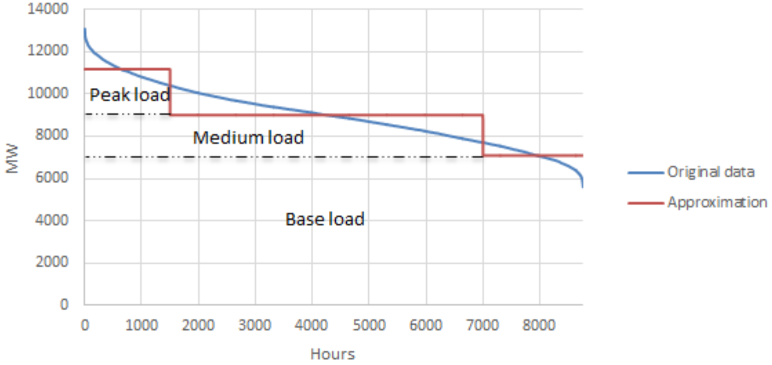
- We know that $V(x_k) = \pi_{k+1}^T(h - Tx_k)$ (why?)
- Since $\theta \geq \pi^T(h - Tx), \pi \in V_k$ is enforced in $(M)$ at iteration $k$, if $V_{k+1} = V_k$ then $\theta_k \geq \pi_{k+1}^T(h - Tx_k)$
- Combining the above relationships, $\theta_k \geq \pi_{k+1}^T(h - Tx_k) = V(x_k)$

Proof that $(S)$ infeasible $\Rightarrow \sigma_{k+1} \notin R_k$

- $\sigma_{k+1}$ is an extreme ray $\Rightarrow \sigma_{k+1}^T(h - Tx_k) > 0$
- If $\sigma_{k+1} \in R_k$, then $\sigma_{k+1}^T(h - Tx_k) \leq 0$ (contradicting the first bullet)
- Therefore, $\sigma_{k+1} \notin R_k$

# Table of Contents

# Load Duration Curve




Load duration curve is obtained by sorting load time series in descending order

# Mathematical Programming Formulation

$$\min_{x,y\geq 0} \sum_{i=1}^{n} (I_i \cdot x_i + \sum_{j=1}^{m} C_i \cdot T_j \cdot y_{ij})$$

$$\text{s.t.} \ \sum_{i=1}^{n} y_{ij} = D_j, j = 1, \ldots, m$$

$$\sum_{j=1}^{m} y_{ij} \leq x_i, i = 1, \ldots n-1$$

- $I_i, C_i$: fixed/variable cost of technology $i$
- $D_j, T_j$: height/width of load block $j$
- $y_{ij}$: capacity of $i$ allocated to $j$
- $x_i$: capacity of $i$

# Problem Data

| Technology | Fuel cost ($/MWh) | Inv cost ($/MWh) |
|---|---|---|
| Coal | 25 | 16 |
| Gas | 80 | 5 |
| Nuclear | 6.5 | 32 |
| Oil | 160 | 2 |

| | Duration (hours) | Level (MW) |
|---|---|---|
| Base load | 8760 | 0-4235 |
| Medium load | 7000 | 4235-7496 |
| Peak load | 1500 | 7496-10401 |

# Benders Decomposition Master

$$(M): \quad \min_{x \geq 0} \sum_{i=1}^{n} l_i \cdot x_i + \theta$$

$$\theta \geq \sum_{j=1}^{m} \lambda_j^v D_j + \sum_{i=1}^{n} \rho_i^v x_i, (\lambda^k, \rho^k) \in V_k$$

$$\theta \geq 0$$

$\lambda_j^k$, $\rho_i^k$: dual optimal multipliers of slave

Note $\theta \geq 0$

- because slave has has non-negative cost
- necessary for boundedness of master

# Benders Decomposition Slave

$$(S): \quad \min_{y \geq 0} \sum_{i=1}^{n} \sum_{j=1}^{m} C_i \cdot T_j \cdot y_{ij}$$

$$(\lambda_j): \quad \sum_{i=1}^{n} y_{ij} = D_j, j = 1, \ldots, m$$

$$(\rho_i): \quad \sum_{j=1}^{m} y_{ij} \leq \bar{x}_i, i = 1, \ldots n-1$$

$\bar{x}_i$: trial decision from master

# Sequence of Investments

| Iteration | Coal (MW) | Gas (MW) | Nuclear (MW) | Oil (MW) |
|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 0 |
| 2 | 0 | 0 | 0 | 8735.6 |
| 3 | 0 | 0 | 0 | 18565.1 |
| 4 | 0 | 14675.8 | 0 | 0 |
| 5 | 10673.3 | 0 | 0 | 0 |
| 6 | 0 | 0 | 7337.9 | 3063.1 |
| 7 | 0 | 1497.7 | 7337.9 | 732.2 |
| 8 | 0 | 1497.7 | 7337.9 | 2033.3 |
| 9 | 0 | 0 | 8966 | 1435 |
| 10 | 2851.8 | 2187.2 | 5362 | 0 |
| 11 | 8321 | 0 | 0 | 2080 |
| 12 | 6989.5 | 4489.5 | 56.5 | 0 |
| 13 | 3261 | 2905 | 4235 | 0 |

# Observations

- A *new* investment proposal is necessarily made in each iteration (why?)
- Greedy behavior
  - First iteration: no investment
  - Early iterations: technologies with low investment cost