Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich
Universität Zürich

Mathematical Optimization, Autumn Semester 2017

Summary of the lectures in 2017

Vanessa Leite

Github (git/svn) repository page:
https://github.com/ssinhaleite/eth-mathematical-optimization-summary
Contact vrcleite@gmail.com if you have any questions.

Friday 26th January, 2018

Professor:
Prof. Dr. Robert Weismantel
D-MATH, ETH, UZH

Summary of the lectures about Mathematical Optimization - Prof. Dr. Robert Weismantel. You can find a referece list in the end of each chapter.

—

Contents

Introduction

Complexity Theory - short revision

Definition: an algorithm is a finite list of instructions. The running time of an algorithm is the number of operations needed to complete it.

Simplifications

- We measure the "complexity" of an algorithm with respect to the length of the input, n. $T_A(n)$ = complexity of the algorithm = worst case for same input size = calculated/defined as the maximum over all inputs of length n.
- We are only interested in how $T_A(n)$ increases w.r.t. n.
  $T_A(n) = O(n)$: "linear"
  $T_A(n) = O(n^k)$: "polynomial"
  $T_A(n) = O(2^n)$: "exponential"
- We assume that every operation of the algorithm requires the same cost ("unit cost model").

Note We need to assume that the number of bits of any stored member must be bounded by a (constant) multiple of the number of bits of the largest input value. Example: if $a \in \mathbb{Z}$ is part of the input, its input size is $O(log|a|)$. An algorithm with running time of $O(a)$ is not polynomial-time! ($O(e^{loga})$: "pseudo polynomial time" $\rightarrow O((logm \times A_{ij}) \times m \times n)$.

Decision Problems

Let $\Sigma = \{0, 1\}$ and Σ^* be a finite binary sequences (sequences on Σ). We assume that all problems are encoded as elements of Σ^*.

Definition $\mathcal{P}$ *problems*: For a set $S \subseteq \Sigma^*$, the decision problem defined by S is the problem wether a given $x \in \Sigma^*$ is in S or not.

It is a polynomial-time solvable ($S \in \mathcal{P}$) if there exists an algorithm $\mathcal{A}$ s.t. $\forall x \in \Sigma^*, \mathcal{A}(x) = 1 \iff x \in S$.

If you can solve a decision problem, usually you can solve the optimization problem.

Definition $\mathcal{NP}$ *problems*: The decision problem defined by $S \subseteq \Sigma^*$ is in $\mathcal{NP}$ if there is a polynomial time algorithm $\mathcal{A} : \Sigma^* \times \Sigma^* \mapsto \{0, 1\}$ and a polynomial p, s.t.:

- $\forall x \in S, \exists y \in \Sigma^*$ s.t. $|y| \in O(p(|x|)$ and $\mathcal{A}(x, y) = 1$
- $\forall x \in S, \forall yin\Sigma^*, \mathcal{A}(x, y) = 0$

Example of $\mathcal{P}$ problems: Let $G = (V, E)$ be an undirected graph. "does G have a cycle?"; "does G have a perfect matching?". Example of $\mathcal{NP}$ problems: "does there exist a cycle covering all vertices (hamiltonian cycle)?". This problem is easy to check, given a cycle, if it covers all vertices, however, it is not that easy find the cycle.

NP-Hard vs NP-Complete A decision problem S is reducible to $S'(S \subseteq pS') \iff$ there is a function $f : \Sigma^* \mapsto \Sigma^*$ compatible in polynomial s.t. $x \in S \iff f(x) \in S'$.

Lemma Let S, S' be decision problems. If $S' \in \mathcal{P}$ (is polynomial-time solvable). Then, $S \subseteq pS' \Rightarrow S \in \mathcal{P}$.

Definition NP-Complete: A decision problem is called NP-complete if $S \in \mathcal{NP}$ and $\forall S' \in \mathcal{NP}, S' \subseteq pS$. If you have a solution for S, you could solve all other S problems in polynomial time.

Definition NP-Hard: A decision problem S is NP-Hard if $\forall S' \in \mathcal{NP}, S' \subseteq pS$.

Optimization problems in finite dimensional space

Definition: Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function. $dom(f) = \{x \in \mathbb{R}^n \mid |f(x)| < \infty\}$.

We need a function to optimize. An optimization problem is of the form:

$$\min_{\mathbf{x} \in \mathcal{F}}/\max \quad f(\mathbf{x})$$

Where $\mathcal{F} \subseteq \mathbb{R}^n$ is the feasible domain given implicitly or by a "membership oracle". The function $\mathcal{F} : \mathbb{R}^n \to \mathbb{R}$ is presented implicitly (i.e., is a domain you can't describe) or by an "evaluation oracle", for instance, given by a query: $x \in \mathbb{R}^n$: true ($x \in \mathcal{F}$) or false ($x \notin \mathcal{F}$).

Implicitly means that you need evaluate by a given point; continuum spectrum of points can be scaled and produce a minimum nore negative than any natural.

What does $\min\limits_{x \in \mathcal{F}} f(x)$ mean for us? This problem can be assigned to one of three meanings:

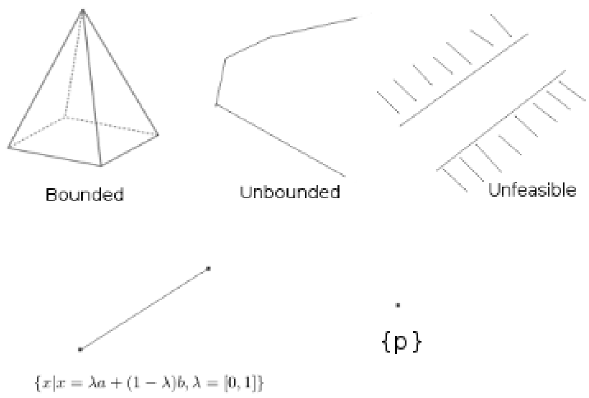
- nothing: the problem is *infeasible*; it means that the domain is empty, i.e., $\mathcal{F} = \emptyset$, there is nothing to optimize.
- the problem is *unbounded*, i.e., there exists a sequence of points $x^1, x^2, \cdots \in \mathcal{F}$ and $\forall n \in \mathbb{N} f(x^i) < -n \Rightarrow lim_{i \to \infty} f(x^i) = -\infty$. There exists a sequence of points that for each natural number, $f(x^i) < -n$ so we cannot bound the sequence.
- the problem has an *optimal solution*, i.e., $\exists x^* \in \mathcal{F}$ such that $f(x^*) \leq f(z) \forall z \in \mathcal{F}$.

Target of the course Derive conclusions (meanings from above 1.2) for optimization problem $\min\limits_{x \in \mathcal{F}} f(x)$ and give ideally a proof that our conclusion is correct. It is necessary to proof (find a proof) to hold the conclusions. What to do when the set $\mathcal{F}$ is empty? It is necessary to use a mathematical theory.

Basic definitions

organize the enumeration of figures in the text

Figure 1: Types of polyhedra.



- A *half space* is a set of the form $\{x \in \mathbb{R}^n \mid a^T x \leq \alpha\}$ for $a \in \mathbb{R}^n, \alpha \in \mathbb{R}$.

  A *rational half space* is a set of the form $\{x \in \mathbb{R}^n \mid a^T x \leq \alpha\}$ for $a \in \mathbb{Q}^n, \alpha \in \mathbb{Q}$. In this case the boundary is a rational line.

- A *polyhedron* is the intersection of a **finite** number of halfspaces (Figure 1.3). The intersection of halfspaces gives a region of finite points. In this case, it is called *polytope*: a bounded polyhedron. We need to use geometry and algebric forms. In algebric form: $\mathcal{P} = \{x \in \mathbb{R}^n \mid Ax \leq b\} = \{x \in \mathbb{R}^n \mid a_1^T x \leq b_1, \ldots, a_n^T x \leq b_n\}$ where $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$.

  Rational polyhedron: $A \in \mathbb{Q}^{m \times n}, b \in \mathbb{Q}^m$.

  one half space is also a polyhedron.

  all points inside a polyhedron are called feasible.

  $Ax \leq b$ represents the multplication of the rows of A with x. This is less equal than a value in same row of the column vector b.

  add a graphical representation of this subitem

- A set $\mathcal{Q}$ is *convex* (Figure 1.3) if $\forall x, y \in \mathcal{Q}$ and $\lambda \in (0,1)$ then $\lambda x + (1-\lambda)y \in \mathcal{Q}$.

- A function f is *linear* if $f(x) = c^T x$ for $c \in \mathbb{R}^n$.

- A function f is *convex* if $f : \mathbb{R}^n \to \mathbb{R} \Rightarrow \emptyset \neq dom(\mathcal{F})$ is a convex set and $f(\lambda x + (1-\lambda)y) \leq \lambda(f(x)) + (1-\lambda)(f(y)) \forall x, y \in dom(\mathcal{F})$ and $\forall \lambda \in (0,1)$.

- A function f is *strictly convex* if $f(\lambda x + (1-\lambda)y) < \lambda(f(x)) + (1-\lambda)(f(y)) \forall x, y \in dom(\mathcal{F})$ and $\forall \lambda \in (0,1)$.

Figure 2: A polyhedron/polytope.

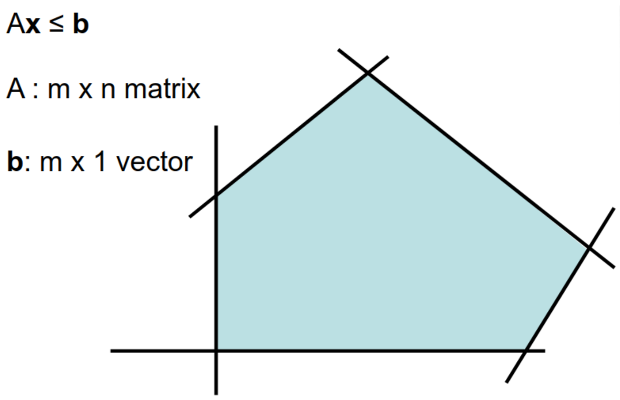
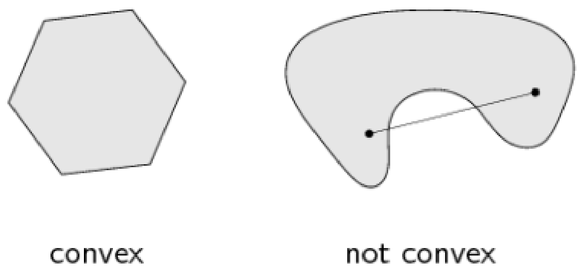


Figure 3: Convex and non-convex set.



- lef f be continuously differentiable, then f is convex if, and only if, $f(y) \geq f(x)+\Delta f(x)^T(yx)\forall y, x \in dom(f)$. And strictly convex if, and only if, $f(y) > f(x)+\Delta f(x)^T(yx)\forall y, x \in dom(f)$ and $y \neq x$.
- If f is twice differentiable, then f is convex if, and only if, $\Delta^2 f(x) \geq 0, \forall x \in dom(f)$ (positive semidefinite).

Convex set Given vectors $(x^1, \ldots, x^t \in \mathbb{R}^n)$ the convex hull of this vectors are defined as $conv(x^1, \ldots, x^t) := \{x \in \mathbb{R}^n \mid \exists \lambda_1, \ldots, \lambda_t \geq 0\}$ such that $\mathbb{1}^T\lambda = 1$ (or $\sum_{i=1}^t \lambda_i = 1$): $x = \sum_{i=1}^t \lambda_i x^i$.

Note $conv(x^1, \ldots, x^t) = \{x \in \mathbb{R}^n \mid x = \sum_{i=1}^t \lambda_i x^i, \lambda \geq 0, \sum \lambda = 1\}$ is a special case of a polyhedron (bounded polyhedron) and every bounded polyhedron is of this form.

Lemma

1. a polyhedron is a convex set
2. $conv(x^1, \ldots, x^t)$ is a convex set

Proof of 1 Let $\mathcal{P} = \{x \in \mathbb{R}^n \mid Ax \leq b\}$. Let $x, y \in \mathcal{P}, \lambda \in (0,1)$. Show $\lambda x + (1-\lambda)y \in \mathcal{P}$.

$$A[\lambda x + (1-\lambda)y] = \lambda Ax + (1-\lambda)Ay$$
$$\lambda Ax \leq \lambda b \text{ and } (1-\lambda)Ay \leq (1-\lambda)b$$
$$\lambda Ax + (1-\lambda)Ay \leq \lambda b + (1-\lambda)b = b$$
$$\Rightarrow \lambda x + (1-\lambda)y \in \mathcal{P}.$$

Proof of 1 Let $x^1, \ldots, x^t \in \mathbb{R}^n$. Let $x, y \in conv(x^1, \ldots, x^t)$ and $0 < \lambda < 1$. Show $\lambda x + (1-\lambda)y \in conv(x^1, \ldots, x^t)$.
x lies in $conv(x^1, \ldots, x^t)$. What does that means? That implies $x = \sum_{i=1}^t \mu_i x^i, \mu_i \geq 0 \forall i$ and $y = \sum_{i=1}^t \sigma_i x^i, \sigma_i \geq 0 \forall i$

$$\lambda x + (1-\lambda)y = \sum_{i=1}^t \lambda \mu_i x^i + \sum_{i=1}^t (1-\lambda)\sigma_i x^i$$
$$= \sum_{i=1}^t (\lambda \mu_i + (1-\lambda)\sigma_i)x^i$$
$$\tau_i = (\lambda \mu_i + (1-\lambda)\sigma_i), \tau_i \geq 0, \forall i$$
$$= \sum_{i=1}^t \tau_i = \sum_{i=1}^t (\lambda \mu_i + (1-\lambda)\sigma_i)$$
$$= \lambda \sum_{i=1}^t \mu_i + (1-\lambda) \sum_{i=1}^t \sigma_i$$
$$\sum_{i=1}^t \mu_i = 1 \text{ and } \sum_{i=1}^t \sigma_i = 1$$
$$= \lambda + (1-\lambda) = 1 \to \sum_{i=1}^t \tau_i x^i \in conv(x^1, \ldots, x^t)$$

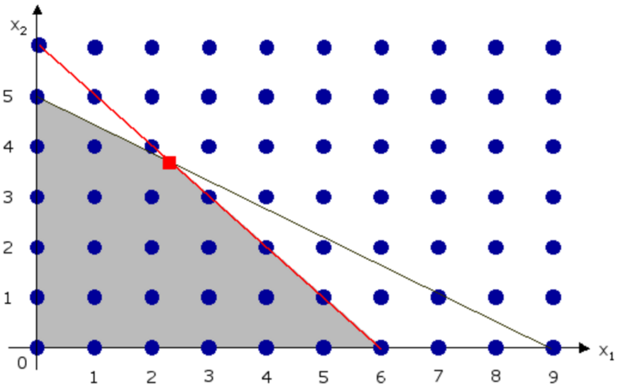
Classification of Optimization Problems

Definition of optimization problem: For $f : \mathbb{R}^n \mapsto \mathbb{R}$ (any function), solve $\min\limits_{x \in \mathcal{F}} f(x)$ (objective value) or $\operatorname*{argmin}\limits_{x \in \mathcal{F}} f(x)$ (value of x that minimize $f(x)$).

Note $\max\limits_{x \in \mathcal{F}} f(x) = \min\limits_{x \in \mathcal{F}} -f(x)$

- *linear optimization problem*: given $A \in \mathcal{Q}^{m \times n}, b \in \mathcal{Q}^m, c \in \mathcal{Q}^n$. $\min\limits_{x \in \mathcal{P}} c^T x$ where $\mathcal{P}$ is a polyhedron defined by $\{x \in \mathcal{R}^n \mid x_i \geq 0 \forall i, Ax = b\}$ or $\max\limits_{x \in \mathcal{P}} c^T x$ where $\mathcal{P} = \{x \in \mathcal{R}^n \mid Ax \leq b\}$
  $c^T x$ is the objective (linear) function.

Figure 4: ILP



$Ax \leq b$ is the constraint/feasible region.
Can be solved quickly and in polinomial time.

integer linear optimization problem (Figure 1.4). $max/min\{c^T x \mid Ax \leq b, x \in \mathbb{Z}^n\}$. Solve a integer optimization problem can lead to a better result than just round linear solutions.

mixed integer linear optimization problem $max/min\{c^T x \mid Ax \leq b, x \in \mathbb{Z}^{n-d} \times \mathbb{R}^d\}$.

- *convex optimization problem*: $\mathcal{F}$ is a convex set, $f : \mathcal{R}^n \to \mathcal{R}$ is a convex function and $\mathcal{F} \subseteq dom(f)$
  $f(x)$ convex and $\mathcal{F}$ convex set. $\forall x, y \in \mathbb{R}^n, \lambda \in [0,1], f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$.
  $\forall x, y \in \mathcal{F}, \lambda x + (1-\lambda)y \in \mathcal{F}$.

  Example of implicitly: knapsack problem: collections are all feasible solutions of knapsack problem.

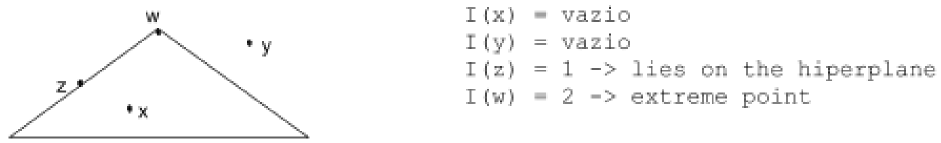
- *combinatorial optimization problem*: given finite groundset $\mathcal{E}$ and a collection $\mathcal{I}$ of subsets of $\mathcal{E}$ (tipically, implicitly given: you cannot write it down). For $c : \mathcal{E} \mapsto \mathcal{R}$ find a member in the collection $(I \in \mathcal{I})$ such that $\sum_{i \in I} c_i$ is minimal/maximal. Example: $G = (V, E), E \subseteq V \times V$ $S \subseteq V$ stable implies that for all pairs in $S(\forall i, j \in S)$, the corresponding edge i, j is not present $(i \neq j, (i,j) \notin E)$. Groundset: V
  $\mathcal{I} \subseteq 2^V$
  $I \in \mathcal{I} \iff I$ is stable in G (implicitly given)
  $c : V \mapsto \mathcal{R}$; suppose $c(v) = 1, \forall c \in V$
  find a maximal (w.r.t. cardinality) stable set in G

- *general integer optimization problem*:
  $\max\limits_{x \in \mathcal{P} \cap \mathbb{Z}^n} c^T x, \mathcal{P} = \{x \in \mathbb{R}^n \mid Ax \leq b\}$ or $\min\limits_{x \in \mathcal{P} \cap \mathbb{Z}^n} c^T x, \mathcal{P} = \{x \in \mathbb{R}^n \mid x_i \geq 0 \forall i, Ax = b\}$

Linear Optimization and Extreme Points I

Geometry can be misleading. We try to get some intuitions from geometry but what we want to do is came up with the algebra.

Consider $\max\limits_{x \in \mathcal{P}} c^T x, P = \{x \in \mathcal{R}^n \mid Ax \leq b\}, A \in \mathbb{Q}^{m \times n}, b \in \mathbb{Q}^m, c \in \mathbb{Q}^n$



From linear algebra we know:
$A \in \mathbb{Q}^{m \times n}$
$n = dim(Ker(A)) + dim(Im(A))$, where $Ker(A)$ is the kernel space and $Im(A)$ is the image space.
$Ker(A) = \{x \in \mathbb{R}^n \mid Ax = 0\}$
$Im(A) = \{x \in \mathbb{R}^n \mid x^T = y^T A, \text{ for some } y \in \mathbb{R}^n\}$
$Ker(A) \perp Im(A)$ (kernel space is orthogonal/perpendicular to image space).
$\forall z \in Im(A)$ and $\forall x \in Ker(A)$ (where z and x are vectors), $z^T x = y^T Ax = 0$ (the dot product is zero).

Definition: Let $P = \{x \in \mathbb{R}^n \mid Ax \leq b\}$ be a polyhedron.

1. $x \in P$ is an extreme point if there is no representation of the form $x = \lambda y + (1-\lambda)z$ for $y, z \in P$, $y \neq z$ and $\lambda \in (0,1)$.
2. $x \in P$ is an vertex if there exists $c \in \mathbb{R}^n$ such that $c^T x > c^T y$, $\forall y \in P \setminus \{x\}$
3. For a point $x \in \mathbb{R}^n$, the index set has tight constraints: $I(x) = \{i \in \{1, ..., m\} \mid A_{i\cdot}x = b_i\}$
4. x is a basic solution if $dim(\{A_{i\cdot} \mid i \in I(x)\} = n$, i.e., the vectors are linearly independent. For a basic feaseble solution, $x \in P$.

Theorem: let $P = \{x \in \mathbb{R}^n \mid Ax \leq b\} \neq \emptyset$. Let $x^* \in P$. The following statements are equivalent:

1. x^* is a vertex
2. x^* is an extreme point
3. x^* is a basic feasible solution

Proof (1 $\rightarrow$ 2): Take a vertex x^*. That means that exist some vector $c \in \mathbb{R}^n$ such that $c^T x^* > c^T x$, $\forall x \in P \setminus \{x^*\}$. Suppose x^* is not an extreme point $\exists y, z \in P$, $y \neq z$ and $\lambda \in (0,1)$ such that $x^* = \lambda y + (1-\lambda)z$. Therefore, $c^T x^* > c^T y$ and $c^T x^* > c^T z$

$$
\begin{aligned}
c^T x^* &= c^T(\lambda y + (1-\lambda)z) \\
&= \lambda c^T y + (1-\lambda) c^T z \\
\lambda c^T y < \lambda c^T x^* and (1-\lambda) c^T z &< (1-\lambda) c^T x^*, thus \\
&< \lambda c^T x^* + (1-\lambda) c^T x^* = c^T x^* \\
c^T x^* &< c^T x^* \to contradiction!
\end{aligned}
$$

Proof (2 → 3) or (¬3 → ¬2): Let $x^* \in P$ (so, it is feasible), assume x^* is not a basic solution, i.e, $dim(\{A_{i.} \mid i \in I(x^*)\}) < n$. Linear algebra tells us $\exists z \in \mathbb{R}^n \setminus \{0\}$ such that $A_{i.}z = 0$, $\forall i \in I(x^*)$.

$$
\begin{aligned}
A_{i.}x^* &= b_i, \forall i \in I(x^*) \\
A_{i.}x^* &< b_i, \forall i \notin I(x^*)
\end{aligned}
$$

Let

$$
\varepsilon = \begin{cases} 1, & \text{if } A_{i.}z = 0, \forall i \notin I(x^*). \\ \min\{\frac{b_i - A_{i.}x^*}{|A_{i.}z|} \mid i \notin I(x^*) \text{ such that } A_{i.}z \neq 0 & \text{otherwise.} \end{cases}
$$

Note $\varepsilon > 0$.

Claim: $y^+ = x^* + \varepsilon z \in P$ and $y^- = x^* - \varepsilon z \in P$. This way, x^* can be written as a convex combination of y^+ and y^-: $x^* = \frac{1}{2}y^+ + \frac{1}{2}y^-$. If x^* is not a basic feasible solution, then it is not an extreme point.

Proof of the claim: wlog, $y^+ \in P$. $\forall i \in I(x^*) : A_{i.}y^+ = A_{i.}x^* + \varepsilon \underbrace{A_{i.}z}_{=0} = b_i$

- $\forall i \notin I(x^*)$ such that $A_{i.}z \leq 0$: $A_{i.}y^+ = A_{i.}x^* + \underbrace{\overbrace{\varepsilon}^{>0} A_{i.}z}_{\leq 0} < b_i$
- $\forall i \notin I(x^*)$ such that $A_{i.}z > 0$: Then $\varepsilon \leq \frac{b_i - A_{i.}x^+}{A_{i.}z}$ and hence, $A_{i.}y^+ = A_{i.}x^* + \varepsilon A_{i.}z \leq A_{i.}x^* + \frac{b_i - A_{i.}x^*}{A_{i.}z} A_{i.}z = b_i$

do the same proof for y^-

Proof (3 → 1): $x^* \in P$ is a basic feasible solution. $dim(\{A_{i.} \mid i \in I(x^*)\}) = n$

By linear algebra: $\{z \in \mathbb{R}^n \mid A_{i.}z = 0, \forall i \in I(x^*)\} = \{0\}$ (*)

Define $c^T = \sum_{i \in I(x^*)} A_{i.} \in \mathbb{R}^n$

Since, $\underbrace{A_{i.}x^* = b_i}_{\text{tight constraints}}$ and $A_{i.}x \stackrel{\leq}{} b_i$:

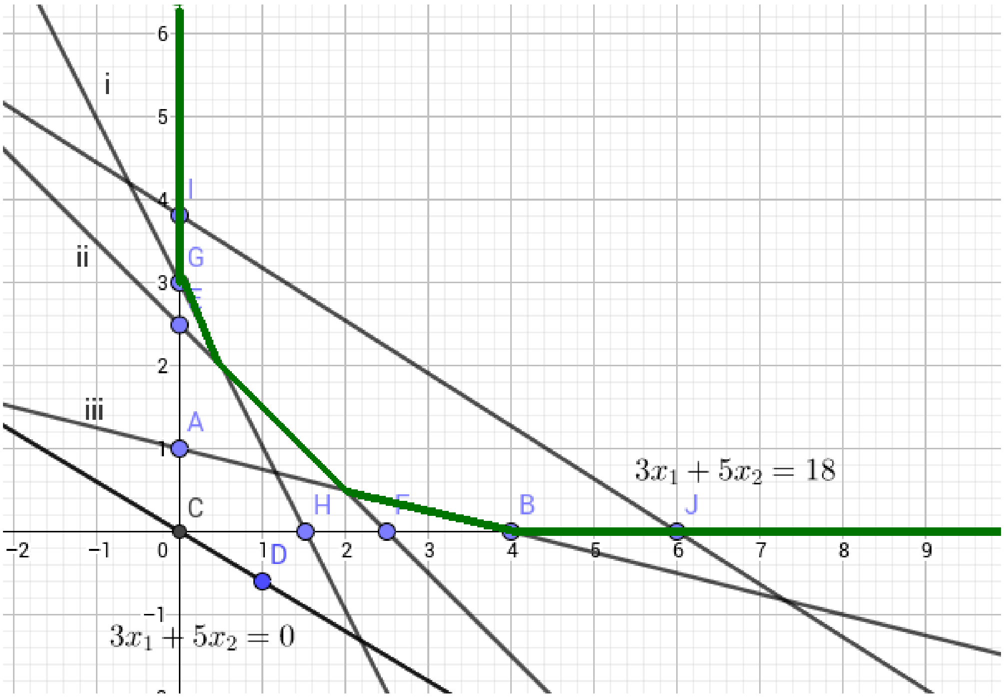
$c^T x^* = \sum_{i \in I(x^*)} A_{i.}x^* = \sum_{i \in I(x^*)} b_i \underbrace{\geq}_{\forall x \in P} \sum_{i \in I(x^*)} A_{i.}x = c^T x$

When is $c^T x^* = c^T x$? Only when $A_i.x^* = b_i \forall i \in I$, i.e, with tight constraints for x. Thus, $A_i.x = A_i.x^*$, $\forall i \in I(x^*)$ $A_i.(x - x^*) = 0$, $\forall i \in I(x^*) \rightarrow$ (*) $x = x^*$. Hence we have that $c^T x^* = c^T x$ implies that $x^* = x$ and $c^T x^* > c^T x \forall x \in P, x \neq x^*$, and thus x^* is a vertex.

Corollary: The number of vertices (extreme points or basic feasible solutions) in a Polyhedron P is finite. **Proof:** $A \in Q^{mxn}$. The number of basic feasible solution is smaller than m^n, and it is a finite number.

Why are extreme points interesting for linear optimization? Consider $\min 3x_1 + 5x_2$

Figure 5: unbounded polytope



$$P = \left\{ x \in \mathbb{R}^2 \mid \begin{array}{c} 2x_1 + x_2 \geq 3 \text{ (i)} \\ 2x_1 + 2x_2 \geq 5 \text{ (ii)} \\ x_1 + 4x_2 \geq 4 \text{ (iii)} \\ x_1, x_2 \geq 0 \end{array} \right\}$$

In the 2D case we can "guess" the right solution. If someone claims that the solution of $\min 3x_1 + 5x_2 = 0$, we can prove it is not by summing up the constraints. From (i) + (iii), we know $\min 3x_1 + 5x_2 \geq 7$. We can find the solution pushing the "lines" until the equation leaves the feasible solution area (marked in green in Figure 2.

Geometrically, $(2, \frac{1}{2})$ is the optimal solution and $(3,5)^T \begin{pmatrix} 2 \\ \frac{1}{2} \end{pmatrix} = 8.5$

Constraint (ii) $\times 1.5 \rightarrow 3x_1 + 3x_2 \geq 7.5$ and $2x_2 \geq 0 \rightarrow 3x_1 + 5x_2 \geq 7.5$. Take (ii) $\times \frac{7}{6}$ and add (iii) multiplied with $\frac{2}{3}$: $3x_1 + 5x_2 \geq 8.5$

More generally, $\min ax_1 + bx_2$:

- Suppose $a < 0$. Take the sequence of points $(k, 0)$ for $k \in \mathbb{N}$, $k \geq 4$
  problem is unbounded
- Suppose $b < 0$. Take the sequence of points $(0, k)$ for $k \in \mathbb{N}$, $k \geq 3$
  problem is unbounded

Linear Optimization and Extreme Points II

Definition: A polyhedron $\mathcal{P}$ contains a line if there exists $d \in \mathbb{R}^n \setminus \{0\}$ and $y \in P$ such that $y + \lambda d \in P \forall \lambda \in \mathbb{R}$

Lemma: Let $\mathcal{P} = \{x \in \mathbb{R}^n \mid Ax \leq b\} \neq 0$, $d \in \mathbb{R}^n \setminus \{0\}$ is a line in $\mathcal{P}$ $\iff$ $Ad = 0$

Proof: If $d \in \mathbb{R}^n \setminus \{0\}$ and $Ad = 0$, then take any $y \in P$, i.e, $Ay \leq b$. Then $\forall \lambda \in \mathbb{R}$, $A(y + \lambda d) = Ay + \underbrace{\lambda \underbrace{Ad}_{=0}}_{=0} \leq b$. Conversely, suppose $d \in \mathbb{R}^n \setminus \{0\}$ is a line in P. Then $\exists y \in P$ such that $y + \lambda d \in P$, $\forall \lambda \in \mathbb{R}$. Suppose, $\exists i \in \{1, \ldots, m\}$ such that $A_i d \neq 0$ wlog: $A_i d > 0$. Let $\lambda^* = \frac{b_i - A_i y}{A_i d}$, for $\lambda > \lambda^*$, then $A_i(y + \lambda d) = A_i y + \lambda A_i d > A_i y + \lambda^* A_i d = A_i y + \frac{b_i - A_i y}{A_i d} A_i d = b_i$

Observation: Suppose I give you a polyhedron non empty $P = \{x \in \mathbb{R}^n \mid Ax \leq b\} \neq 0$ and $m < n$ then P contains a line!

Theorem: Let $P = \{x \in \mathbb{R}^n \mid Ax \leq b\} \neq 0$. P has an extreme point iff P does not contain a line.

Proof: "$\leftarrow$" P does not contain a line then P has an extreme point Let $x \in P$ such that $l = dim(\{A_{i\cdot} \mid i \in I(x)\})$. If $l = n$ (maximum), then x is a basic feasible solution, thus, an extreme point. Suppose $l < n$. Then from linear algebra, there is a vector $d \in \mathbb{R}^n \setminus \{0\}$ such that $A_{i\cdot}.d = 0$, $\forall i \in I(x)$. d is not a line $\rightarrow$ $\exists j \notin I(x)$ such that $A_{j\cdot}.d \neq 0$. wlog: $A_{j\cdot}.d \geq 0$. Let $J = \{j \notin I(x)\} \mid A_{j\cdot}.d > 0\} \neq 0$.

Notice that for $j \in J$, $A_{j\cdot}$ is linearly independent from $\{A_{i\cdot} \mid i \in I(x)\}$.

Linearly independent: Suppose $A_j = \sum_{i \in I(x)} A_i \lambda_i$, then $\underbrace{A_{j\cdot}.d}_{=0} = \sum_{i \in I(x)} \lambda_i A_{i\cdot}.d = 0$. So, it is linearly independent.

Let, $\lambda^* = \min\{\frac{b_j - A_{j\cdot}}{A_{j\cdot}.d} \mid j \in J\}$. Then $x + \lambda^* d \in P$. Moreover, we observe $A_{i\cdot}.(x + \lambda^* d) = b_i$, $\forall i \in I(x)$. Let $j \in J$ such that $\underbrace{\lambda^* = \frac{b_j - A_{j\cdot}.x}{A_{j\cdot}.d}}_{\text{the minimum}}$, then $A_{j\cdot}.(x + \lambda^* d) = A_{j\cdot}.x + \frac{b_j - A_{j\cdot}.x}{A_{j\cdot}.d} A_{j\cdot}.d = b_j$.

$A_{j\cdot}$ is linearly independent from $\{A_i \mid i \in I(x)\} \rightarrow dim(\{A_{k\cdot} \mid k \in I(x+\lambda^* d)\}) = l+1$, so, this is a contradiction of l be maximal.

Proof: "$\rightarrow$" P has an extreme point then P does not contain a line Let x^* be an extreme point, then $dim(\{A_{i\cdot} \mid i \in I(x^*)\}) = n$. Then, $A_{i\cdot}d = 0$, $\forall i \in I(x^*)$ implies $d = 0$. Therefore, $Ad = 0$ implies $d = 0$, by using the previous lemma, there is no line in P.

Lemma: optimization problem $\rightarrow$ maximize a linear function let $P = \{x \in \mathbb{R}^n \mid Ax \leq b\}$ and $c \in \mathbb{R}^n$. Suppose that i) P has an extreme point and ii) there exists an optimal solution to $\max\limits_{x \in P} c^T x$. Then, there exists an extreme point solution attaining the optimal solution.

Proof: Let $v^* = \max\limits_{x\in P} c^T x =$ optimal value. Let $Q = \{x \in P \mid c^T x = v^*\} = \{x \in \mathbb{R}^n \mid Ax \leq b, c^T x \leq v^*, -c^T x \leq -v^*\}$. Q is a polyhedron and $Q \neq 0$ (there is an optimal solution). $Q \subseteq P$, since P has an extreme point. P has no line $\rightarrow$ Q has no line $\rightarrow$ Q has an extreme point x^*. Claim: x^* is an extreme point in P

Proof of the claim: Suppose $\exists y, z \in P$, $y \neq z$ and $\lambda \in (0,1)$ such that $x^* = \lambda y + (1-\lambda)z$. $v^* = c^T x^* = \lambda c^T y + (1-\lambda) c^T z \leq \lambda c^T x^* + (1-\lambda) c^T x^* = c^T x^* \rightarrow \underbrace{c^T y = v^* \text{and } c^T z = v^*}_{\text{to have an equality in previous equation}} \rightarrow$ $y, z \in Q$, x^* is an extreme point in Q.

Theorem: Let $P \neq 0$ be a polyhedron not containing a line. Then, $\max\limits_{x\in P} c^T x$ is either equal to $+\infty$ or there exists an extreme point in P attaining the optimal value.

Proof: We must show that if the optimal value is not infinite, then there exists an optimal solution. In fact, we proof a stronger claim.

Claim: if the max value is not infinite then, for every $x \in P$, there exists an extreme point, $w \in P$, such that $c^T x \leq c^T w$. From the claim, the statement follows: Let $\{w^1, \ldots, w^r\}$ be all extreme points in P, not empty, also r is finite. Let $\underbrace{w}_{\text{is the maximum}} \in \{w^1, \ldots, w^r\}$ allows $max\{c^T w^1, \ldots, c^T w^r\}$. Then, $\forall x \in P$, $\exists w^i$ such that $c^T x \leq c^T w^i$, but by definition, $\leq c^T w$, then, w is an optimal solution.

Proof of the claim: Let $x^* \in P$ and $dim(\{A_{i\cdot} \mid i \in I(x^*)\}) = k < n$. If it were equals to n, we could use the point itself. $\exists d \in \mathbb{R}^n \setminus \{0\}$ such that $A_{i\cdot}d = 0$, $\forall i \in I(x^*)$. wlog: $c^T d > 0$.

- if $A_{j\cdot} \leq 0$, $\forall j \in I(x^*)$, then

  $x^* + \lambda d \in P$, $\forall \lambda \geq 0$. If max value is not infinite then, $c^T d = 0$. $\exists j \in I(x^*)$ such that $A_{j\cdot}d < 0$ (otherwise, d is a line). We observe $A_{j\cdot}$ is linearly independent from the constraints $\{A_{i\cdot} \mid i \in I(x^*)\}$. Let $\lambda^* = min\{\frac{b_j - A_{j\cdot}x^*}{|A_{j\cdot}d|} \mid A_{j\cdot}d < 0\}$. Then, following in the negative direction, $x^* - \lambda^* d \in P$ and $c^T(x^* - \lambda^* d) = c^T x^* - \lambda^* c^T d = c^T x^*$. Then, $dim(\{A_{i\cdot} \mid i \in I(x^* - \lambda^* d)\} = k+1$. And now we can iterate with $x^* - \lambda^* d$ in place of x^*.

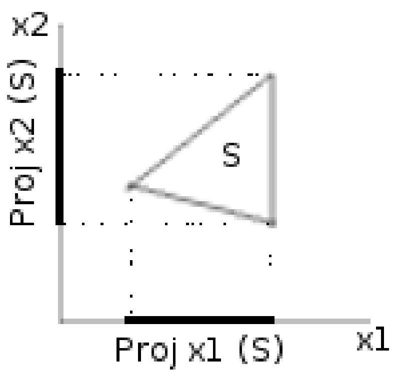
- if there exists $j \notin I(x^*)$ such that $A_{j}.d > 0$

  $A_{j}.$ is linearly independent from $\{A_{i}. \mid i \in I(x^*)\}$ and then $\lambda^* = min\{\frac{b_j - A_i.x^*}{A_i.d} \mid A_j.d > 0\}$ and $x^* + \lambda^* d \in P$ and $dim(\{A_{i}. \mid i \in I(x^*)\}) = k+1$ and we can iterate with $x^* + \lambda^* d$ in place of x^*.

Projection of Polyhedra

Elimination of one variable

Definition: Let $S \subseteq \mathbb{R}^n$**.** $Proj_{x_1}, \ldots, x_{n-k}(S) = \{(x_1, \ldots, x_{n-k}) \in \mathbb{R}^{n-k} \mid \exists x_{n-k+1}, \ldots, x_n$ **such that** $(x_1, \ldots, x_{n-k-1}, x_{n-k}, \ldots, x_n) \in S\}$**.** See figure 13.1.



Notation: Let $P = \{x \in \mathbb{R}^n \mid Ax \leq b\} = \{x \in \mathbb{R}^n \mid \sum_{j=1}^{n} a_{ij} x_j \leq b_i,\ \forall i \in M\}$

For $x \in \mathbb{R}^n$, $\bar{x} = \begin{pmatrix} x_1 \\ \ldots \\ x_{n-1} \end{pmatrix}$ (we drop the last variable). $\bar{A} = [A_{\cdot 1}| \ldots |A_{\cdot (n-1)}]$

Algorithm to compute projection ($proj_{\bar{x}}(P)$**)**

01. Rewrite the system $Ax \leq b$ **in another way: by isolating one variable and normalizing.** $Ax \leq b = a_{in} x_n \leq b_i - \sum_{j=1}^{n-1} a_{ij} x_j \forall i \in M$. Then, we obtain a new representation: $x_n \leq d_i + f_i^T \bar{x} \forall i$ such that $a_{in} > 0$, $x_n \geq d_i + f_i^T \bar{x} \forall i$ such that $a_{in} < 0$, $0 \leq d_k + f_k^T \bar{x} \forall k$ such that $a_{kn} = 0$. where, $d_i = \frac{b_i}{a_{in}}$ and $f_i = -(\frac{a_{ij}}{a_{in}}) \forall j = 1, \ldots, n-1$.

02. Define $Q = \{\bar{x} \in \mathbb{R}^{n-1} \mid d_j + f_j^T \bar{x} \leq d_i + f_i^T \bar{x} \forall j, i$ **such that** $a_{in} > 0$**,** $a_{jn} < 0$ **and** $0 \leq d_k + f_k^T \bar{x} \forall k$ **such that** $a_{kn} = 0\}$ Standard form: $f_j^T \bar{x} - f_i^T x \leq d_i - d_j$

In the worst case, the number of constraints generated by the algorithm goes to m^2**, considering** m **is the original number of constraints.**

03. Return Q Theorem: $Q = proj_{\bar{x}}(P)$ (we verify the number of constraints in Q match the dimension)

Proof: "$proj_{\bar{x}}(P) \subseteq Q$" Let $\bar{x} \in proj_{\bar{x}}(P) \to \exists x_n$ such that $(\bar{x}, x_n) \in P$. Then $(\bar{x}, x_n)$ satiesfies the new representation in the first step of the algorithm. Thus, $\bar{x} \in Q$.

Proof: "$Q \subseteq proj_{\bar{x}}(P)$" Let $\bar{x} \in Q$. Let $L := max\{d_j + f_j^T \bar{x} \mid a_{jn} < 0\}$ and $U := min\{d_i + f_i^T \bar{x} \mid a_{in} > 0\}$. $L \leq U$, since $\bar{x} \in Q$. Take any $x_n \in [L, U]$. Then $(\bar{x}, x_n)$ satisfies the new representation in the first step of the algorithm. Thus, $\bar{x} \in proj_{\bar{x}}(P)$.

Projection is a key algebraic operator for duality

The special case of dimension 1

Take a polyhedron in dimension 1 (an interval): $P \subseteq \mathbb{R}$. Let $a \in (\mathbb{Q} \setminus \{0\})^m$, $b \in \mathbb{Q}^m$. $P = \{x \in \mathbb{R} \mid ax \leq b\}$, we can rewrite P as: $P = \{x \in \mathbb{R} \mid l \leq x \leq u\}$, $l = max\{\frac{b_i}{a_i} \mid a_i < 0\}$ and $u = min\{\frac{b_i}{a_i} \mid a_i > 0\}$.

When the polyhedron is empty $(P = \emptyset)$? $P = \emptyset \iff l > u \iff 0 > u - l$. This conclusion can be derived from the original representation of P.

Using a projection $proj_0(P) = \{0 \leq y^T b \; \forall y \geq 0,\, y^T a = 0\}$
Observation: $P \neq \emptyset \iff 0 \leq y^T b \; \forall y \geq 0$ such that $y^T a = 0$.

Repeated projections

Observation: $proj_{(x_1,\ldots,x_{n-2})}(P) = proj_{(x_1,\ldots,x_{n-2})} proj_{(x_1,\ldots,x_{n-1})}(P))$.

Theorem: Let $P = P^{(0)} = \{x \in \mathbb{R}^n \mid Ax \leq b\}$. Define a submatrix $A^{(j)}$ of matrix A with columns $A_{\cdot k} \forall k \in \{1, \ldots, j\}$. Also define a set $C^{(0)} = \mathbb{R}^m_+$. $C^{(i)} = \{y \in \mathbb{R}^m_+ \mid y^T A_{\cdot k} = 0$ $\forall k = n-i+1, \ldots, n\} = \{y \in C^{(i-1)} \mid y^T A_{\cdot(n-i+1)} = 0\}$

Then $proj_{(x_1,\ldots,x_{n-i})}(P)$ can be described by half-spaces: $P^{(i)} = \{\bar{x} \in \mathbb{R}^{n-i} \mid y^T A^{(n-i)} x \leq y^T b$ $\forall y \in C^{(i)}\}$.

Proof: "$proj_{(x_1,\ldots,x_{n-i})} \subseteq P^{(i)}$" Take $(x_1, \ldots, x_{n-i}) \in proj_{x_1,\ldots,x_{n-i}}(P)$. $\exists z \in \mathbb{R}^j$ such that $\begin{pmatrix} x_1 \\ \cdots \\ x_{n-1} \\ z \end{pmatrix} \in P$.

Take any $y \in C^{(i)}$: Show $y^T A^{(n-i)} x \leq y^T b$, $y \geq 0$. $\sum_{j=1}^{n-i} A_{.j} x_j + \sum_{j=n-i+1}^{n} A_{.j} z_j \leq b \rightarrow \sum_{j=1}^{n-i} y^T A_{.j} x_j + \underbrace{\sum_{j=n-i+1}^{n} y^T A_{.j} z_j}_{=0,\text{ because } y \in C^{(i)}} \leq y^T b$.

$\iff y^T A^{(n-i)} \leq y^T b$.

Proof: "$P^{(i)} \subseteq proj_{(x_1,\ldots,x_{n-i})}(P)$" We perform induction on i.
$i = 0 \rightarrow$ trivial.
Suppose our theorem applies for indices up to $j-1$. Consider index j.
What we know: the $proj_{(x_1,\ldots,x_{n-j})}(P) = proj_{(x_1,\ldots,x_{n-j})}(proj_{(x_1,\ldots,x_{n-j+1})}(P)$.

Inductive argument: $proj_{(x_1,\ldots,x_{n-j})}(P) = P^{(j-1)} = \underbrace{\{x \mid y^T A^{(n-j+1)} x \leq y^T b \forall y \in C^{(j-1)}\}}_{\text{infinite, but redundat}} = \{x \in \mathbb{R}^{n-j+1} \mid \underbrace{Bx \leq d}_{\text{finite constraints}}\}$.

Define:
$M_+ = \{i \in M \mid B_{i,n-j+1} > 0\}$
$M_- = \{i \in M \mid B_{i,n-j+1} < 0\}$
$M_0 = \{i \in M \mid B_{i,n-j+1} = 0\}$

Then, $proj_{(x_1,\ldots,x_{n-j})}(P^{(j-1)}) = Q \overbrace{=}^{\text{first theorem}} \{\bar{x} \in \mathbb{R}^{n-j} \mid (a) \sum_{k=1}^{n-j} B_{j,k} \bar{x}_k \leq d_i \forall i \in M_0,\ (b) \sum_{k=1}^{n-j} (\frac{B_{r,k}}{B_{r,n-j+1}} - \frac{B_{s,k}}{B_{s,n-j+1}}) \bar{x}_k \leq \frac{d_r}{B_{r,n-j+1}} - \frac{d_s}{B_{s,n-j+1}},\ \forall r \in M_+,\ s \in M_-\}$.

Claim: Every constraint of type (a), (b) in Q is a constraint in $P^{(i)}$.
Take a constraint of type (a) in Q. We know $i \in M_0$ and the induction implies $\exists y \in C^{(j-1)}$ such that $y^T A^{(n-j+1)} = B_i$. Since $B_{i,n-j+1} = 0 \rightarrow y \in C^{(j)}$.

Take a constraint of type (b) in Q. Inductive argument shows $\exists y^r \in C^{(j-1)}$, $y^s \in C^{(j-1)}$ such that $(y^r)^T A^{n-j+1} = B_r$, $(y^s)^T A^{n-j+1} = B_s$.
$y = \frac{1}{B_{r,n-j+1}} y^r - \frac{1}{B_{s,n-j+1}} y^s \in C^{(j)} \rightarrow y^T A_{.(n-j+1)} = 0$ and $y \geq 0$, $y \in C^{(j-1)} \rightarrow y \in C^{(j)}$.

Farkas Lemma and Standard Form Polyhedra

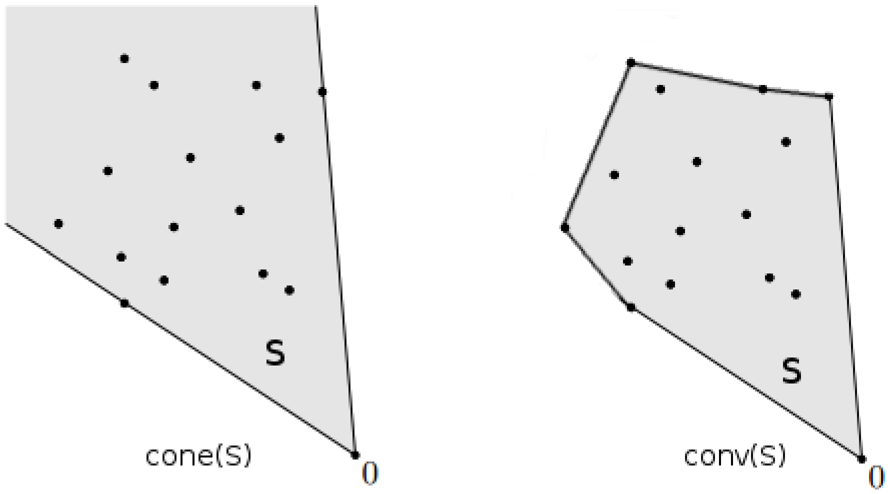
Some consequences from lecture 4

Corollary: If I take a polyhedron $P = \{x \in \mathbb{R}^n \mid Ax \leq b\}$, the projection $proj_{(x_1,\ldots,x_i)}(P)$ is a polyhedron (by removing points).

Corollary: Take a polyhedron $P = \{x \in \mathbb{R}^n \mid Ax \leq b\}$, consider the set Q (the set of all vectors of the form: $\{wx \mid x \in P\}$, where $w \in \mathbb{Q}^{d \times n}$; $wx \subseteq \mathbb{R}^d$. Q is a polyhedron.

Proof: Let's define another set $D = \{(x, y) : wx - y = 0,\ x \in P\} = \{(x, y) \mid wx - y = 0, Ax \leq b\} \subseteq \mathbb{R}^{n+d}$. Q is the projection of D on the space of $Q = proj_y(D) = \{y \in \mathbb{R}^d \mid \exists x \in \mathbb{R}^{d\times n}$ such that $\underbrace{(x, y) \in D}_{\text{that means } wx = y}\}$.

Corollary: Let $x^1, \ldots, x^t \in \mathbb{Q}^n$ (set of vectors). $conv(x^1, \ldots, x^t)$ is a polyhedron.



Corollary: Let $x^1, \ldots, x^t \in \mathbb{Q}^n$ (set of vectors). $cone(x^1, \ldots, x^t)$ is a polyhedron. Define a matrix $X = [x^1 \mid x^2 \mid \cdots \mid x^t] \in \mathbb{Q}^{n\times t}$.

Proof: convex hull $conv(x^1, \ldots, x^t) = \{\sum_{i=1}^t \lambda_i x^i \mid \lambda_i \geq 0 \forall i, \sum_{i=1}^t \lambda_i = 1\}$.
$= \{X \times \lambda \mid \lambda \in \mathbb{R}^t_+,\ \lambda_i \geq 0,\ \mathbb{1}^T\lambda = 1\} \rightarrow$ is a polyhedron from the previous corollary.

Proof: cone Cone is unbounded and with a finite many rational vectors, a rational cone is a polyhedron. $cone(x^1, \ldots, x^t) = \{\sum_{i=1}^t \lambda_i x^i \mid \lambda_i \geq 0 \forall i\}$. $= \{X \times \lambda \mid \lambda \in \mathbb{R}^t_+,\ \lambda_i \geq 0\} \rightarrow$ is a polyhedron from the previous corollary.

Corollary: Take a polyhedron $P \subseteq \mathbb{R}^n$. Then $P \neq \emptyset \iff proj_{(x_1,\ldots,x_{n-1})}(P) \neq \emptyset$

Theorem - Farka's Lemma: Either $Ax \leq b, x \in \mathbb{R}^n$ has a solution or $y^T A = 0, y^T b < 0, y \geq 0$ has a solution. Both are exclusive. If $y^T A = 0$ and $y^T b < 0, y \geq 0$ then $Ax \leq b, x \in \mathbb{R}^n$ is empty.

Proof - do not exist simultaneously x and y such that both systems ae feasible Suppose this is not true, then $\underbrace{y^TAx}_{=0} \underbrace{\leq}_{y\geq 0} y^Tb \to 0 \leq y^Tb < 0$.

Let $P = \{x \in \mathbb{R}^n \mid Ax \leq b\}$. For $L \in \{1, \dots, n\}$, let $C^{(i)} = \{y \geq 0 \mid y^TA_{\cdot k} = 0, \forall k \geq n - i + 1\}$. $P^{(i)} = proj_{(x_1,\dots,x_{n-i})}(P) = \{\bar{x} \in \mathbb{R}^{n-i} \mid \sum_{k=1}^{n-i} y^TA_{\cdot k}\bar{x}_k \leq y^Tb \forall y \in C^{(i)}\}$.

Then $C^{(n)} = \{y \geq 0 \mid y^TA = 0\}$, $P^{(n)} = \{0 \leq y^Tb \forall y \in C^n\}$ $P \neq \emptyset \iff P^{(1)} \neq \emptyset \iff \cdots \iff P^{(n)} \neq \emptyset$. Hence, either $P \neq \emptyset$ or $P = \emptyset \iff P^{(n)} = \emptyset \iff \exists y \geq 0$ such that $y^TA = 0$ and $y^Tb < 0$.

Standard form polyhedra

Definition: Take a matrix $A \in \mathbb{Q}^{m \times n}$ of full row rank. Then take a vector $b \in \mathbb{Q}^n$. A polyhedron of the type $\{x \in \mathbb{R}^n_+ \mid Ax = b\}$ is called in standard form.

Observation: $\{x \in \mathbb{R}^n_+ \mid Ax = b\}$ is a polyhedron.

A polyhedron of the form $(\{x \in \mathbb{R}^n \mid Ax \leq b\}$ can be brought into standard form:

$Ax \leq b, x \in \mathbb{R}^n \to$ split into positive and negative i.e., For $x \in \mathbb{R}^n$ write $x = x^+ - x^-, x^+, x^- \geq 0$.

Artificially transform in equations by introducing slack variables (how far we are from the plane): $A_{i\cdot}x \leq b_i \to A_{i\cdot}x + y_i = b_i, y_i \geq 0$.

Then, we can represent $\{x \in \mathbb{R}^n \mid Ax \leq b\}$ as $\{(x^+, x^-, y) \in \mathbb{R}^{2n+m} \mid [A| - A| \underbrace{I}_{\text{for the y's}}] \begin{pmatrix} x^+ \\ x^- \\ y \end{pmatrix} = b, x^+, x^-, y \geq 0\}$.

Observation: A polyhedron in standard form has no lines. $P = \{x \in \mathbb{R}^n_+ \mid Ax = b\} \subseteq \mathbb{R}^n_+$.

- theorem from lecture 3 applies.
- optima are characterized by extreme points, if the optima is finite.

Theorem: $Ax = b, x \geq 0$ has a solution iff $\forall y \in \mathbb{R}^n$ such that $y^TA \geq 0$ then $y^Tb \geq 0$.

Proof proof as homework!

Theorem: $x^* \in P = \{x \in \mathbb{R}^n_+ \mid Ax = b$ (polyhedron in standard form (= full row rank))$\}$ is a basic feasible solution iff there exists a basis $B \subseteq \{1, \dots, n\}$, $|B| = m$ (number of constraints), with the properties:

- $A_B \in \mathbb{R}^{m \times m}$ is invertible
- $x^*_i = (A_B^{-1}b)_i$: $\forall i \in B$
- $x^*_i = 0$, $\forall i \in N = \{1, \dots, n\} \setminus B$

Theorem: Let $v^1, \dots, v^t \in \mathbb{Z}^n$ be vectors and $y \in cone(v^1, \dots, v^t)$. There exists a subset $B \subseteq \{1, \dots, t\}$, $|B| \leq n$ and multipliers $\mu_i > 0 \forall i \in B$ such that $y = \sum_{i \in B} \mu_i v^i$

Proof: Consider $c^* = min \sum_{i=1}^{t} \lambda_i$, st $V\lambda = y, \lambda \geq 0$, $V = [v^1| \dots |v^t] \in \mathbb{Z}^{m \times t}$. $Q = \{\lambda \in \mathbb{R}^t_+ \mid V\lambda = y\}$ is a standard form polyhedron. From previous arguments, $c^* \geq 0 \rightarrow c^*$ is finite, i.e., there exists extreme point $x^* \in Q$ attaining the optimum solution. $\rightarrow \exists B \subseteq \{1, \dots, t\}, |B| = n$, such that $x^*_i = 0, \forall i \notin B$.

Theorem of separating hiperplane (consequence of Farka's Lemma): For $v^1, \dots, v^t \in \mathbb{Z}^n$, let $b \in \mathbb{R}^n$. $b \in cone(v^1, \dots, v^t)$ or $\exists c$ (a hiperplane) $\in \mathbb{R}^n$, such that $c^T x \geq 0 \forall x \in cone(v^1, \dots, v^t)$ and $c^T b < 0$.

Proof: Let $V = [v^1, \dots, v^t] \in \mathbb{Z}^{n \times t}$. b is not in the cone ($b \notin cone(v^1, \dots, v^t)$) iff our linear system ($V\lambda = b, \lambda \geq 0$) has no solution.

Farka's lemma $\rightarrow \exists y \in \mathbb{R}^n$ such that $y^T V \geq 0, y^T b < 0$.
Define $c = y$.

- $c^T b = y^T b < 0$
- $\forall x \in cone(v^1, \dots, v^t), x = V\lambda, \lambda \geq 0$

$c^T x = \underbrace{c^T V}_{\geq 0} \underbrace{\lambda}_{\geq 0} \geq 0$.

Linear Programming Duality

Theorem (Neuman '47): Let $P = \{x \in \mathbb{R}^n \mid Ax \leq b\} \neq \emptyset$. Assume there exists a point in the polyhedron. $D = \{y \in \mathbb{R}^m \mid y \geq 0,\ y^T A = c^T\} \neq \emptyset$. $\max\limits_{x \in \mathcal{P}} c^T x = \min\limits_{y \in D} y^T b$

Proof: D is a polyhedron in standard form, $D \neq \emptyset$, so, there is no line, then, D has an extreme point.
Since $\mathcal{P} \neq \emptyset$, take any $x \in \mathcal{P}$. $c^T x \underbrace{=}_{\forall y \in \mathcal{D}} \underbrace{y^T}_{\geq 0} \underbrace{Ax}_{\leq b} \leq y^T b$, therefore, the minimum is not infinity.
$\exists y^* \in \mathcal{D}$, extreme point such that $\underbrace{\delta^*}_{\text{optimal value}} = y^{*^T} b = min\{y^T b \mid y \in \mathcal{D}\}$
$\forall x \in \mathcal{P}$, then $c^T x = y^{*^T} Ax \leq \delta^* \rightarrow \max\limits_{x \in \mathcal{P}} c^T x$ is bounded (can not go to $+\infty$).

This way,

$$\max_{x \in \mathcal{P}} c^T x \leq \min_{y \in \mathcal{D}} y^T b$$

Now, consider all points in $\mathcal{P}$ such that $c^T x \geq \delta^*$. This set must be non empty, because for some x, $c^T x = \delta^*$.
$\{x \in \mathcal{P} \mid c^T x \geq \delta^*\} \neq \emptyset$. Suppose it is empty, then:
$Ax \leq b$ has no solution and $-c^T x \leq -\delta^*$ (multiplied by -1, so, every inequality is organized in the same way).

By Farka's Lemma, there exists multipliers $z \in \mathbb{R}^m_+$ and $\lambda \in \mathbb{R}_+$ (z is a multiplier for the first inequality and λ for the second).

$$\begin{pmatrix} z \\ \lambda \end{pmatrix} \begin{pmatrix} A \\ -c^T x \end{pmatrix} = 0$$

and

$$\begin{pmatrix} z \\ \lambda \end{pmatrix} \begin{pmatrix} b \\ -\delta^* \end{pmatrix} < 0$$

so, $z^T A - \lambda c^T = 0$, $z^T b - \lambda \delta^* < 0$.
Assume $\lambda = 0$:
By Farka's Lemma, $\exists$ solution $z \geq 0$ such that $z^T A = 0$ and $z^T b < 0 \iff \mathcal{P} = \emptyset$. This is a contradiction by definition.
Assume $\lambda > 0$:
Since $\lambda > 0$ we can divide the system by λ, this way define $\bar{y} = \frac{z}{\lambda}$. Then, $z^T A = \lambda c^T \iff \bar{y}^T A = c^T$, $z^T b < \lambda \delta^* \iff \bar{y}^T b < \delta^*$.
δ^* is $\min\limits_{y \in \mathcal{D}} y^T b$, so it is a contradiction $y^T b < \delta^* \rightarrow \{x \in \mathcal{P} \mid c^T x \geq \delta^*\} \neq \emptyset$, otherwise we would have found another point better than the minimum.

This way, $\exists x \in \mathcal{P}$ such that $c^T x \geq \delta^*$. Since $\forall x \in \mathcal{P}$, $c^T x \leq \delta^*$, we can conclude that $c^T x = \delta^*$, thus,

$$\max_{x \in \mathcal{P}} c^T x = \min_{y \in \mathcal{D}} y^T b$$

Weak complementary slackness

Theorem: $\mathcal{P} = \{x \in \mathbb{R}^n \mid Ax \leq b\} \neq \emptyset$ **and** $\mathcal{D} = \{y \geq 0 \mid y^T A = c^T\} \neq \emptyset$**. We say** $x \in \mathcal{P}$ **and** $y \in \mathcal{D}$ **are simultaneously optimal solutions iff** $y_i(A_i.x - b_i) = 0 \ \forall i = 1, \ldots, m$**.**

Proof:

$$c^T x = y^T A x = \sum_{i=1}^{m} y_i A_i.x = \sum_{i:y_i>0} y_i A_i.x \leq \sum_{i:y_i>0} y_i b_i = y^T b$$

$x \in \mathcal{P}, y \in \mathcal{D}$ are simultaneously optimal $\underbrace{\text{iff}}_{duality}$ $c^T x = y^T b \iff \forall i : y_i > 0$, then $A_i x = b_i \iff y_i(A_i.x - b_i) = 0$.

Observation: What happens when $\mathcal{P}$ or $\mathcal{D}$ are empty?

Lemma: assume $\mathcal{P}$ is non empty, but the dual $\mathcal{D}$ is empty. $\mathcal{P} = \{x \in \mathbb{R}^n \mid Ax \leq b\} \neq \emptyset$, $D = \{y \in \mathbb{R}^m_+ \mid \underbrace{y^T A = c^T}_{=A^T y = c, y \geq 0}\} = \emptyset$. Then $\max_{x \in \mathcal{P}} c^T x = +\infty$.

Proof: $\mathcal{P} \neq \emptyset \to \exists \bar{x} \in \mathcal{P}$. $\mathcal{D} = \emptyset \underbrace{\to}_{\text{farka's lemma}} \exists z$ such that $Az \geq 0, c^T z \leq 0$.

Claim: $\bar{x} - \lambda z \in \mathcal{P}, \forall \lambda \geq 0$
$A(\bar{x} - \lambda z) = A\bar{x} - \lambda \underbrace{Az}_{\geq 0} \leq b$
$c^T(\bar{x} - \lambda z) = c^T \bar{x} - \lambda c^T z \xrightarrow[x \to \infty]{} +\infty$.

Lemma: assume $\mathcal{D}$ is non empty, but the primal $\mathcal{P}$ is empty. $\mathcal{D} = \{y \in \mathbb{R}^m_+ \mid y^T A = c^T\} \neq \emptyset$, $\mathcal{P} = \{x \in \mathbb{R}^n \mid Ax \leq b\} = \emptyset$. Then $\min_{y \in \mathcal{D}} y^T b = -\infty$.

Proof: $\mathcal{D} \neq \emptyset \to \exists \bar{y} \in \mathcal{D}$. $\mathcal{P} = \emptyset \underbrace{\to}_{\text{farka's lemma}} \exists y : y^T A = 0, y^T b < 0 \to y < 0$.

Claim: $\bar{y} + \underbrace{\lambda y}_{\text{you can scale infinitely}} \in \mathcal{D}$, $\forall \lambda \geq 0$ and the sequence of points $b^T[\bar{y} + \lambda y] \xrightarrow[\lambda \to \infty]{} -\infty$.

A short summary

	$\mathcal{P} = \emptyset$	$\mathcal{P} \neq \emptyset$
$\mathcal{D} = \emptyset$	Both are infeasible	$\max_{x \in \mathcal{P}} c^T x = +\infty$
$\mathcal{D} \neq \emptyset$	$\min_{y \in \mathcal{D}} y^T b = -\infty$	$\max_{x \in \mathcal{P}} c^T x = \min_{y \in \mathcal{D}} y^T b$

Dualization operator

If we agree on a linear optimization of the following form: $\max_{Ax \leq b} c^T x$, we can create a "dualization" operator.

$$Dual(\max_{x \in \mathcal{P}} c^T x) = \min_{y^T A = c^T, y \geq 0} y^T b$$

Observation: The dual of the dual linear programming is the primal, provided all are non empty.

By applying the dual, we first "drop" the sign on the objective function:
$\min_{y \in \mathcal{D}} y^T b = \max -y^T b$, $A^T y \leq c$, $-A^T y \leq -c$, $-Iy \leq 0$

$Dual(\max(-b)^T y),\ A^T y \leq c,\ -A^T y \leq -c,\ -Iy \leq 0) = \min -c^T \underbrace{(v-u)}_{=z},\ A(v-u)+w = b,$
$u, v, w \geq 0 = \min -c^T z,\ Az + w = b,\ w \geq 0$

Then, we "reapply" the sign on the objective function $\min -c^T z,\ Az \leq b = \max c^T z,\ Az \leq b$

Supplementary Reading - Section 4.3 from [BT97]

update

Representation of Polyhedra

Definition: Let $\mathcal{P} = \{x \in \mathbb{R}^n \mid Ax \leq b\}$ be a polyhedron. $dim(\mathcal{P}) = n - dim(\{A_{i\cdot} \mid A_{i\cdot}x = b_i, \forall x \in \mathcal{P}\})$ (implicitly linearly independent equations).

For $I \subseteq \{1, \ldots, m\}$, $F = \{x \in \mathcal{P} \mid A_{i\cdot}x = b, \forall i \in I\}$ is a face of $\mathcal{P}$. Extreme case: $F = \emptyset$, $F = P$.

Vertices are faces of dimension 0. Edges are faces of dimension 1. Facets are faces of dimension $dim(\mathcal{P}) - 1$.

When $b = 0$, $\mathcal{P}$ is a cone. A cone is **pointed** if it does not contain a line.

For $\mathcal{P}$ and a point $y \in \mathcal{P}$, $\underbrace{rec}_{\text{recession}}(y, \mathcal{P}) = \{d \in \mathbb{R}^n \mid y + \lambda d \in \mathcal{P}, \forall \lambda \geq 0\}$. If $\mathcal{P}$ is bounded then $rec(y, \mathcal{P}) = \{0\}$, $\forall y \in \mathcal{P}$.

Observation: if $\mathcal{P} = \{x \in \mathbb{R}^n \mid Ax \leq b\}$, then the recession cone $rec(y, \mathcal{P}) = \{d \in \mathbb{R}^n \mid Ad \leq 0\}$, $\forall y \in \mathcal{P}$. Therefore for every point $y \in \mathcal{P} \rightarrow rec(y, \mathcal{P}) = rec(\mathcal{P})$. In other words, the recession cone is independent of y.

Proof of observation: Take $y \in \mathcal{P}$, $rec(y, \mathcal{P}) = \{d \in \mathbb{R}^n \mid y + \lambda d \in \mathcal{P}, \forall \lambda \geq 0\}$, now we claim this is the same as $\{d \in \mathbb{R}^n \mid Ad \leq 0\} = \mathcal{Q}$.

"$Q \subseteq rec(y, \mathcal{P})$": if $d \in \mathcal{Q} \rightarrow Ad \leq 0 \underbrace{\rightarrow}_{\forall y \in \mathcal{P}} A(y + \lambda d) = Ay + \underbrace{\lambda}_{\geq 0} \underbrace{Ad}_{\leq 0} \leq b$.

"$rec(y, \mathcal{P}) \subseteq Q$": Suppose it is not true, so, there exists i, such that for $d \in rec(y, \mathcal{P}) \rightarrow A_{i\cdot}d > 0$. $y + \lambda d \in \mathcal{P}$, $\forall \lambda \geq 0 \rightarrow A_{i\cdot}(y + \lambda d) = A_{i\cdot}y + \lambda A_{i\cdot}d \leq b_i \rightarrow \lambda \leq \frac{b_i - A_{i\cdot}y}{A_{i\cdot}d}$, which contradicts $\lambda > 0$.

Remark: For a polyhedron in standard form $P = \{x \in \mathbb{R}^n_+ \mid Ax = b\}$, $rec(\mathcal{P}) = \{d \in \mathbb{R}^n_+ \mid Ad = 0\}$.

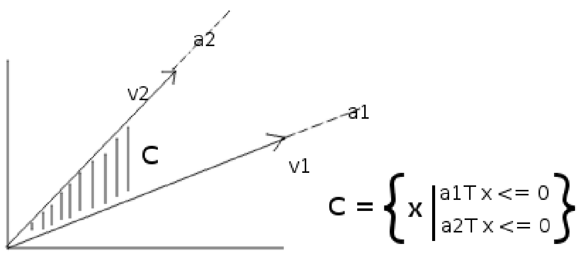
From lecture 3, we know that if we take a matrix $A \in \mathbb{Z}^{m \times n}$ and a cone $C = \{x \in \mathbb{R}^n \mid Ax \leq 0\}$. Then $(a) \iff (b) \iff (c)$:

- (a) 0 is an extreme point
- (b) C is pointed
- (c) the dimension of all inequalities should be n $(dim(\{A_{i.} \mid i = 1, \cdots, m\}) = n)$.

A cone can not have more than one extreme point. In fact, for all $x \in C$ if $x \neq 0 \to x = \frac{1}{2}[\frac{1}{2}x + \frac{3}{2}x]$ (there is no linear combination to represent x (?) - it seems that x can be written as a linear combination).

check above information!

Definition: Let the cone $C = \{x \in \mathbb{R}^n \mid Ax \leq 0\}$. $x \in C$, $x \neq 0$ is called an extreme ray of C if the dimension of all the tight constraints at this point is $n-1$ $(dim(\{A_{i.} \mid i \in I(x)) = n-1)$. For $\mathcal{P} = \{x \in \mathbb{R}^n \mid Ax \leq b\}$ an extreme ray of $\mathcal{P}$ is an extreme point of $rec(\mathcal{P})$.



Remark: Extreme rays of C are faces of C with dimension 1. From now on, if $A \in Q^{m \times n}$, extreme rays are scaled to be integers whose $gcd = 1$. $C = \{x \mid x = \lambda_1 v_1 + \lambda_2 v_2, \lambda_1, \lambda_2 \geq 0\}$ v_1 and v_2 are the extreme rays of C.
Can you give a point $c^T x \geq 0$? If v_1 and v_2 are ≥ 0 then yes. Otherwise, we have a point that is false.

Theorem: Take a polyhedron $P = \{x \in \mathbb{R}^n \mid Ax \leq b\}$ with at least one extreme point. What happens if $\max\limits_{x \in \mathcal{P}} c^T x = +\infty$? This can happen iff there is an extreme ray $d \in \mathcal{P}$, such that $c^T d > 0$.

Proof: Suppose d extreme ray of $\mathcal{P}$, $c^T d > 0$. $P \neq \emptyset$. Take a point $x \in \mathcal{P}$. By definition of extreme rays, $Ad \leq 0$, hence $x + \lambda d \in \mathcal{P}$, $\forall \lambda \geq 0$.
$\to c^T(x + \lambda d) = c^T x + \lambda c^T d \xrightarrow[\lambda \to \infty]{} +\infty$.

Conversely, $\max\limits_{x \in \mathcal{P}} c^T x = +\infty$, shows that exists an extreme ray.

homework

Consider now that the dual $\min b^T y$, $st.A^T y = c$, $y \geq 0$ has no solution. We know that $\min 0^T y$, $st.A^T y = c$, $y \geq 0$ is infeasible. Consider now the cone problem: $\max c^T x$, $st.Ax \leq 0$. This problem is either unbounded or infeasible. 0 is feasible for our problem, so, $\max c^T x,\ st. Ax \leq 0$ is unbounded. As $\max c^T x$ is unbounded, $\max c^T x = +\infty$.

$\mathcal{P}$ has at least one extreme point, therefore has no lines, and in particular, $dim(\{A_{i\cdot} \mid i = 1, \ldots, m\}) = n$. If $\mathcal{P}$ does not have a line, then the cone does not have a line, i.e, $rec(\mathcal{P})$ is pointed. Under this assumption in the cone, prove that exists a recession ray in the cone such that $c^T d > 0$.

homework: Start by taking any point of the polyhedron (check lecture 03)

Theorem: Let $\mathcal{P} = \{x \in \mathbb{R}^n \mid Ax \leq b\}$ with at least one extreme point. Let $\{v^1, \ldots, v^k\}$ be all the extreme points in $\mathcal{P}$ and $\{w^1, \ldots, w^r\}$ be all the extreme rays in $\mathcal{P}$.
Then $\mathcal{P} = \mathcal{Q} = \{\sum_{i=1}^k \lambda_i v^i + \sum_{j=1}^r \mu_j w^j \mid \lambda \geq 0,\ \mu \geq 0,\ 1^T\lambda = 1\}$ In other words: every point of a bounded polyhedron can be writed as a convex combination of the vertices and extreme rays of the polyhedron.

Proof: "$\mathcal{Q} \subseteq \mathcal{P}$" and "$\mathcal{P} \subseteq \mathcal{Q}$"

"$\mathcal{Q} \subseteq \mathcal{P}$"
Take an arbitrary point $x \in \mathcal{Q}$. $x = \sum_{i=1}^k \lambda_i v^i + \sum_{j=1}^r \mu_j w^j$, $\lambda \geq 0$, $\mu \geq 0$. We know $Aw^j \leq 0,\ \forall j$, $Av^i \leq b,\ \forall i$. $\mathcal{P}$ convex $\rightarrow \sum_{i=1}^k \lambda_i v^i = y \in \mathcal{P}$, $\lambda_i \geq 0$, $1^T\lambda = 1$. $\rightarrow z = \sum_{j=1}^r \mu_j w^j$ satisfies $Az \leq 0$. $x = y + z$ satisfies $Ax = \underbrace{Ay}_{\leq b} + \underbrace{Az}_{\leq 0} \leq b \rightarrow x \in \mathcal{P}$.

"$\mathcal{P} \subseteq \mathcal{Q}$"
Assume the statement is not true. This implies there exists $z \in \mathcal{P}$ and $z \notin \mathcal{Q}$. Then, $R = \min \sum_{i=1}^k 0\lambda_i + \sum_{j=1}^r 0\mu_j,\ st. \sum_{i=1}^k \lambda_i v^i + \sum_{j=1}^r \mu_j w^j = z,\ \lambda_i,\ \mu_j \geq 0,\ 1^T\lambda = 1$. If $z \notin \mathcal{Q}$, R is infeasible, so the dual problem is unbounded. The dual of R is $\max p^T z + q,\ st. p^T v^i + q \leq 0$, $p^T w^j \leq 0,\ \forall i = 1, \ldots, k,\ \forall j = 1, \ldots, r$.

Which means $p = 0$ and $q = 0$ is feasible $\rightarrow R$ is unbounded. $\rightarrow \exists (p^*, q^*)$ such that $(p^*)^T z + q^* > 0 \geq (p^*)^T v^i + q^*, \forall i = 1, \ldots, k$, $(p^*)^T w^j \leq 0,\ \forall j = 1, \ldots, r$.

$\rightarrow (p^*)^T w^j \leq 0,\ \forall j = 1, \ldots, r,\ (p^*)^T v^i < (p^*)^T z,\ \forall i = 1, \ldots, k.$(*)

Consider $\max (p^*)^T x,\ st. Ax \leq b$, which is feasible ($\mathcal{P} \neq \emptyset$).

Suppose, $\max\limits_{x \in \mathcal{P}} (p^*)^T x = +\infty$, then we know from previous theorem that exists an extreme ray j such that $(p^*)^T w^j > 0$, but this contradicts (*).

Suppose, $\max\limits_{x \in \mathcal{P}} (p^*)^T x$ is finite, i.e, there exists an extreme point v^i , $i \in \{1, \ldots, k\}$ attaining the optimal solution. Since $z \in \mathcal{P}$, then $(p^*)^T v^i \geq (p^*)^T z$, which contradicts (*).

So, this contradicts that $\exists z \in \mathcal{P},\ z \notin \mathcal{Q}$.

Remarks

A pointed polyhedron can be written as the convex hull of its vertices plus the convex hull of its extreme rays ($\mathcal{P} = conv(V) + cone(E)$), where V are the vertices and E the extreme rays of $\mathcal{P}$.

What if $\mathcal{P}$ does contain a line? e.g. $\{x \in \mathbb{R}^2 \mid x_2 \geq 1\} = \underbrace{\begin{pmatrix} 0 \\ 1 \end{pmatrix}}_{\text{convex hull of vertices, just one vertex in this case}} + cone(\left\{\begin{matrix} 0 \\ 1 \end{matrix}\right\}) + cone(\{\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \end{pmatrix}\})$

This idea can be generalized: Let $\mathcal{P} = \{x \in \mathbb{R}^n \mid Ax \leq b\}$ for $A \in \mathbb{Q}^{m \times n}$, $b \in \mathbb{Q}^m$ be a polyhedron containing a line.

Let $d = dim(ker(A))$, $L \in \mathbb{Q}^{n \times d}$ be a matrix whose column vectors form a basis of $ker(A)$, and $Q \in \mathbb{Q}^{n \times n-d}$ be a matrix st. its columns form a basis for $Im(A^T)$. Recall from linear algebra that if we get the two basis together we form a basis for the entire space. $[L|Q]$ forms a basis of $\mathbb{R}^n$ and that $\forall \mu \in \mathbb{R}^{n-d} \setminus \{0\}$. $AQ\mu \neq 0$.

$\mathcal{P} = \{L\lambda + Q\mu \mid \lambda \in \mathbb{R}^d,\ \mu \in \mathbb{R}^{n-d},\ Ay \leq b\}$ = this is a Mikowski's sum $= \{L\lambda \mid \lambda \in \mathbb{R}^d\} + \underbrace{\{y \in \mathbb{R}^n \mid y = Q\mu, \mu \in \mathbb{R}^{n-d}, Ay \leq b\}}_{=\mathcal{P}'}$.

$\mathcal{P}'$ does not contain a line. Let $d \in \mathbb{R}^n$ st $d = Q\alpha$, $\alpha \in \mathbb{R}^{n-d}$, and $Ad = 0$, then, $d = 0$, so, indeed, $\mathcal{P}$ does not contain a line.

Corollary of representation theorem: Let $\mathcal{P}$ be a polyhedron, $\mathcal{P} = \{x \in \mathbb{R}^n \mid Ax \leq b\}$. Then, exist finite sets V and E such that $P = conv(V) + cone(E) + lin(L) = conv(V) + cone(E, L, -L)$. $lin(L) = \{x \in \mathbb{R}^n \mid x = L\lambda,\ \lambda \in \mathbb{R}^d\}$.

Supplementary Reading - Section 4.8 and 4.9 from [BT97]

update

Simplex Algorithm I

$A \in \mathbb{Q}^{m \times n}$, $b \in \mathbb{Q}^n$. Consider $\min\{c^T x \mid Ax = b,\ x \geq 0\}(\mathcal{P})$. and $max\{b^T y \mid y^T A \leq c^T\}(D)$.

We assume $m < n$ (otherwise, extra rows are linearly dependent, and for $m = n$ we have exact one solution or no solution).

Recall: $B \subseteq \{1, \ldots, n\}$ is a basis if $|B| = m$ and A_B is invertible. A_B submatrix of $A \in \mathbb{R}^{m \times n}$ with column vectors $A_{\cdot,j}$ for $j \in B$.

Notation For $x \in \mathbb{R}^n$ we write $x = \begin{pmatrix} x_B \\ x_N \end{pmatrix}$, where $N = \{1, \ldots, n\} \setminus B$, i.e, N is the complement of B.

Definition: Given a basis B, we define the following points: $(y^*)^T = c_B^T A_B^{-1}$ and $x^* = \begin{pmatrix} x_B^* \\ x_N^* \end{pmatrix}$ with $x_N^* = 0$, $x_B^* = A_B^{-1} b$

y^* is a dual basic solution (not necessarily feasible), the same way, x^* is a primal basic solution (not necessarily feasible), and they satisfy complementary slackness: $\forall i$: $x_i^*(c_i - (y_j^*)^T A_{\cdot j}) = 0$.

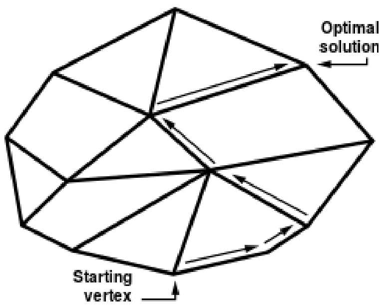
Lemma:

1. y^* is a dual feasible $\iff \forall j \in N$: $c_j - c_B^T A_B^{-1} A_{\cdot j} \geq 0$
2. x^* is a primal feasible $\iff A_B^{-1} b \geq 0$

Proof:

1. y^* is dual feasible $\iff (y^*)^T A \leq c^T \iff \forall j$: $(y^*)^T A_{\cdot j} \leq c_j \iff \forall j \in N$: $c_B^T A_B^{-1} A_{\cdot j} \leq c_j \iff \forall j \in N$: $c_j - c_B^T A_B^{-1} A_{\cdot j} \geq 0$.
2. $Ax^* = A_B x_B^* + A_N \underbrace{x_N^*}_{=0} = A_B x_B^*$, x^* is primal feasible $\iff x^* \geq 0$ and $\forall i \in N$: $x_i = 0 \iff 0 \leq x_B^* = A_B^{-1} b$

Definition: Reduced cost We define the vector of reduced cost $\bar{c} \in \mathbb{R}^{n-m}$ by $\bar{c_j} = c_j - c_B^T A_B^{-1} A_{\cdot j}$
Define $\bar{b} = A_B^{-1} b$
Define $\bar{A} = A_B^{-1} A = [\mathcal{I} | A_B^{-1} A_N]$
The basic idea for an algorithm:



Take initial point $x_o \in \mathcal{P}$. While x_0 not optimal, determine direction $d \in \mathbb{R}^n$, $\lambda > 0$ st $c^T d \leq 0$ and $x_0 + \lambda d \in \mathcal{P}$. Set $x_1 = x_0 + \lambda d$ and iterate.

Questions: How do we find an initial point? How to determine d and λ fast? When to stop?

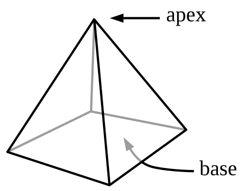
Degeneracy

Definition:

1. For $\mathcal{P} = \{x \in \mathbb{R}^n \mid Ax \leq b\}$ (canonical form), a bfs x^* is degenerated if $|I(x^*)| > n$, where $I(x^*) = \{i \in \{1, \dots, m\} \mid A_{i\cdot}x^* = b_i\}$, i.e, more than n constraints are tight.
2. For $\mathcal{P} = \{x \in \mathbb{R}^n_+ \mid Ax = b\}$ (standard form), a bfs x^* is degenerated if more than $n - m$ components of x^* are equal to zero, i.e, some entries of x^*_B are zero.

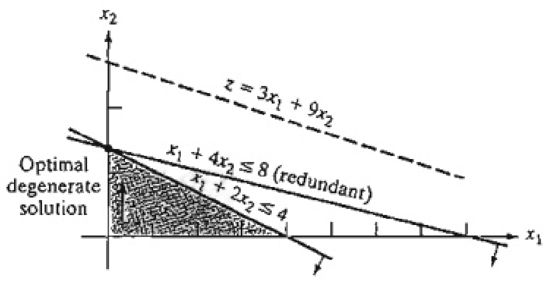
There are differents ways such that degeneracy arises.
Examples:

(i)



The vertex at the apex of a pyramid has four tight constraints, but lies in three dimensions. This is a geometric artifact, and it is the worst case for a simplex.

(ii)



In a triangle, each vertex has two tight constraints. Adding a new inequality is redundant (do not change the polyhedron). This is a redundant artifact, and it is easy to get rid off in the simplex.

Note: In dimension 2, degeneracy $\iff$ redundant. Geometric artifacts starts at dimension 3.

A constraint $A_k x \leq b_k$ is redundant if $\{x \in \mathbb{R}^n \mid Ax \leq b\} = \{x \in \mathbb{R}^n \mid A_1 x \leq b_1, \dots, A_{k-1}x \leq b_{k-1}, A_{k+1}x \leq b_{k+1}, \dots, A_m x \leq b_m\}$.

(iii)
$\{x \in \mathbb{R}^3 \mid \begin{bmatrix} 1 & -1 & 0 \\ 1 & 1 & 1 \end{bmatrix} x = \begin{bmatrix} 0 \\ 2 \end{bmatrix},\ x \geq 0\}$

BFS: $\begin{pmatrix}1\\1\\0\end{pmatrix}$ for $B = \{1, 2\}$ or $\begin{pmatrix}0\\0\\2\end{pmatrix}$ for $B = \{1, 3\}, \{2, 3\}$

Theorem: Let x^*, y^* as defined previously, x^* is feasible, then:

1. if $\bar{c} \geq 0 \rightarrow x^*$ is optimal.
2. if x^* is optimal and non degenerated $\rightarrow \bar{c} \geq 0$
3. if y^* is feasible for the dual and $A_B^{-1}b \geq 0$ then y^* is optimal

Proof:

1. Show that for any other point, the objective value is worst.
   Let $z \in \mathcal{P}$, $d = z - x^* \rightarrow Ad = 0$.
   Write $Ad = A_B d_B + A_N d_N$. Now, $Ad = 0 \iff d_B = -A_B^{-1} A_N d_N$.
   Consider $c^T d = c_B^T d_B + c_N^T d_N = c_N^T d_N - c_B^T A_B^{-1} A_N d_N = \sum_{j \in N} (c_j - c_B^T A_B^{-1} A_{\cdot j}) d_j = \sum_{j \in N} \bar{c}_j d_j$.
   As $z \in \mathcal{P} \rightarrow \forall j \in N$, $z_j = \underbrace{x_j^*}_{=0} + d_j \geq 0 \rightarrow d_N \geq 0$.
   But then, $c^T z = c^T x^* + c^T d = c^T x^* + \sum_{j \in N} \underbrace{\bar{c}_j}_{\geq 0} \underbrace{d_j}_{\geq 0} \geq c^T x^*$. So, $c^T x^*$ is, indeed, minimal.
2. Let x^* be non-degenerate $\rightarrow x_i^* > 0$, $\forall i \in B$. Assume that $\exists j \in N$: $\bar{c}_j < 0$ and x^* is optimal.
   Let $d = \begin{pmatrix} d_B \\ d_N \end{pmatrix} = \begin{pmatrix} -A_B^{-1} A_{\cdot j} \\ \underbrace{e_j}_{(0,\ldots,0,1,0,\ldots) \in \mathbb{Z}^{|N|}} \end{pmatrix}$. Now, $\exists \varepsilon > 0$ st $x^* + \varepsilon d \in \mathcal{P}$.
   $Ad = -\underbrace{A_B A_B^{-1}}_{Im} + \underbrace{A_N e_j}_{A_{\cdot j}} = 0.$
   As $x_i^* > 0$, $\forall i \in B$, $x_i^* + \varepsilon d_i \geq 0$. Now, $c^T(x^* + \varepsilon d) = c^T x^* + \varepsilon c^T d = c^T x^* + \varepsilon \bar{c}_j \underbrace{d_j}_{=1} =$
   $c^T x^* + \varepsilon \underbrace{\bar{c}_j}_{<0} < c^T x^*$.
   So, by contradiction, no other point improves objective value of x^*.
3. Let $z \in \mathcal{D}$ feasible for the dual.
   By definition, $z^T A \leq c^T \rightarrow z^T A_B \leq c_B^T$. Now, $(y^*)^T b = c_B^T \underbrace{A_B^{-1} b}_{\geq 0} \geq z^T A_B A_B^{-1} b = z^T b$. So, y^* is optimal.

Simplex Algorithm II

Consider $\min\{c^T x \mid Ax = b, x \geq 0\}$ as the primal and $\max\{b^T y \mid A^T y \leq c\}$ as the dual.

What is a basis? A subset of the variables with size N. A subset of the columns with size N, they are rows linearly independent, i.e., $A \in \mathbb{R}^{m \times n}$ (full row rank), B basis, $B \subseteq \{1, \ldots, m\}$, $|B| = n$.

A_B^{-1} **exists.** Given a basis it can be associated with two points $B \rightarrow \begin{pmatrix} x^* \\ y^* \end{pmatrix}$. They can be feasible or not. $(x_B^*, x_N^*) = (A_B^{-1}b, 0)$ and $y_B^* = c_B^T A_B^{-1}$.

B is optimal basis if $\bar{b} \geq 0$ and $\bar{c} \geq 0$.

Simplex algorithm is jumping between basis, geometrically jumping between vertices. This is jumping basis because, every time we change the selected vertex, we change the variables of the basis.

Definition: Given $x^* \in \mathcal{P}$, $d \in \mathbb{R}^n$. d is a feasible direction if we can move along the vector d, i.e, if there exists $\lambda > 0$ such that $x^* + \lambda d \in \mathcal{P}$. λ needs to be greather than zero to move far from x^*, but this can be a problem in degeneracy.

Definition: $x^* \in \mathcal{P}$, that is a basic feasible solution $\rightarrow \exists$ basis B such that $x^* = (x_B^*, x_N^*) = (A_B^{-1}, 0)$. The j-th non-basis direction is the vector $d^* \in \mathbb{R}^n$ (uniquely determined), such that for $j \in N$, $d_N^* = \underbrace{e_j \in \mathbb{R}^{|N|}}_{\text{all zeros but the } j\text{-th position}}$ (non-basis variables), $d_B^* = -A_B^{-1} A_{\cdot j}$.

Lemma: Given a bfs x^* for the primal. Assume that x^* is non-degenerate ($\bar{b}$ is strictly positive in all components). The j-th non-basis direction is feasible.

Proof: Let d^* be the j-th non-basis direction. $\underbrace{Ad^* = 0}_{Ax^*=b, \text{ stays at the same kernel}} = A_B d_B^* + A_N d_N^* = -A_B A_B^{-1} A_{\cdot j} + A_N e_j = -A_{\cdot j} + A_{\cdot j} = 0$.

$x^* + \lambda d^* \in \mathcal{P}$, $\forall \lambda \geq 0 \rightarrow x^* + \lambda d^* \geq 0$ such that $\lambda \leq \frac{x_i^*}{-d_i^*}$, $\forall i \in B$ such that $d_i^* < 0$.

In the non-basis, $x_i^* + \lambda d^*$ is always positive because d is the unit vector. So, take $i \in B$:

$$x_i^* + \lambda d_i^* = \begin{cases} \geq x_i^* \geq 0, \ \forall i \in B \text{ such that } d_i \geq 0 \\ \geq x_i^* + \frac{x_i^*}{-d_i^*} d_i^* = 0, \ \forall i \in B \text{ such that } d_i < 0. \end{cases}$$

Observation: the j-th non-basis direction d^* satisfies $c^T d^* = \bar{c_j}$.

$$c^T d^* = c_B^T d_B^* + c_N^T d_N^* = -c_B^T A_B^{-1} A_{\cdot j} + c_j = \bar{c_j}$$

Theorem: Let x^* be a bfs for the primal, with associated basis B. Let $j \notin B$ st $\bar{c_j} < 0$. Assume x^* is non-degenerate. Let d be the j-th non-basis direction:

1. if $d_i \geq 0 \ \forall i \in B \rightarrow \min_{x \in \mathcal{P}} c^T x = -\infty$.

2. if $d \not\geq 0$ (some components are smaller than zero), then $\lambda^* = \min\{\frac{x_i^*}{-d_i} \mid i \in B \text{ st } d_i < 0\} = \max\{\lambda \geq 0 \mid x^* + \lambda d \in \mathcal{P}\}$.

Proof:

1. $Ad = 0 \to A(x^* + \lambda d) = b$, $\forall \lambda \geq 0$. $c^T d = \bar{c}_j < 0$. x^* is feasible so, $x^* \geq 0$. If $d \geq 0$ then $x^* + \lambda d \geq 0$, $\lambda \geq 0$. Then $x^* + \lambda d \in \mathcal{P}$, $\forall \lambda > 0$, $c^T(x^* + \lambda d) = c^T x^* + \lambda \bar{c}_j \xrightarrow[\lambda \to \infty]{} -\infty \to \min_{x \in \mathcal{P}} c^T x = -\infty$.

2. $Ad = 0$, $x_i^* + \lambda d_i \geq 0$, $\lambda \geq 0 \to \lambda \leq \frac{x_i^*}{-d_i}$, $\forall i \in B$, such that $d_i < 0$, $\to \lambda^* = \max\{\lambda > 0 \mid x^* + \lambda d \in \mathcal{P}\}$.

Theorem: Let x^* be a bfs of $\mathcal{P}$ (which is not necessarily non-degenerate). Suppose that there exists an index $j \in N$ st $\bar{c}_j < 0$. Let d be the j-th non basis direction. Assume that at least one entry of d is strictly negative ($d < 0$). Let $k \in B$, be an index reaching the minimum in the definition of λ^*, i.e., $\lambda^* = \frac{x_k^*}{-d_k} = \min_{i \in \{1,\ldots,n\}} \frac{x_i}{-d_i}$. Let $\bar{B} = B \setminus \{k\} \cup \{j\}$:

1. $\bar{B}$ is a basis of x^* w.r.t. to $\mathcal{P}$.
2. $x^* + \lambda^* d$ is the bfs associated with $\bar{B}$.

Proof:

1. Proof of $\bar{B}$ be a basis:

   $i \in \{1, \ldots, n\} \to$ true

   $|B| = n \to$ true.

   Now, show A_B^{-1} exists (and has full row rank).

   A_B is non singular $\iff$ $A_B^{-1} A_{\bar{B}}$ is non singular.
   Let d be the j-th non-basis direction. $d_B = -A_B^{-1} A_{\cdot j}$, $d_k \leq 0$, so, $A_B^{-1} A_{\bar{B}}$ is non singular, because $(A_B^{-1} A_{\cdot j})_k \neq 0$.

2. From previous proof, $\bar{B}$ is a basis. Define $y = x^* + \lambda^* d$. $\forall i \in N \setminus \{j\}$: $y_i = x_i^* + \lambda^* \underbrace{d_i}_{d_i = 0;\ d_j = 1} = 0 \to y_k = x_k^* + \lambda^* d_k = x_k^* + \frac{x_k^*}{-d_k} d_k = 0$.

We have found a feasible solution y to the system: $Az = b$, $z_i = 0$ $\underbrace{\forall i \notin \bar{B}}_{i \in N \setminus \{j\} \cup \{k\}}$. This system has a unique solution: $\hat{z}$. $\hat{z}_B = A_{\bar{B}}^{-1} b$, $\hat{z}_N = 0$. $\hat{z} = y$.

Simplex Algorithm - primal setting

1. Choose a bfs x^* with basis B and cost $c^* = c^T x^*$.
2. Compute the reduced cost: $\bar{c_j} = c_j - c_B^T A_B^{-1} A_{\cdot j}$, $\forall j \in N$.
3. if $\bar{c} \geq 0$, stop $\rightarrow$ optimal!
4. Otherwise, determine $j \in N$, such that $\bar{c_j} < 0$.
   if $d_B = -A_B^{-1} A_{\cdot j} \geq 0$, stop $\rightarrow$ unbounded!
5. Otherwise, compute $\lambda^* = \min_{i \in B} \{ \frac{x_i^*}{-d_i} \mid i \in B, d_i < 0 \}$ and index $k \in B$ such that $\lambda^* = \frac{x_k^*}{-d_k}$.
6. Update the basis and the solutions:
   $B = B \setminus \{k\} \cup \{j\}$ and set $x_k^* = 0$
   $x_i^* = x_i^* + \lambda^* d_i$, $\forall i \in B \setminus \{k\}$
   $x_j^* = \lambda^*$
   $c^* = c^* + \lambda^* \bar{c}_j$

Theorem: For $\mathcal{P} \neq \emptyset$ and assuming that all basic feasible solutions in $\mathcal{P}$ are non degenerate, then the simplex algorithm terminates with a correct answer in finite time.

Proof: Either the minimum value is $-\infty$ or the minimum value is finite. Every pivot operation decreases the objective function value and leads us to a new basis (finitely many), and no basis is visited twice.

Simplex Algorithm III

The worst-case running time of a simplex algorithm

Assumption: we select a pivot-column $j \in N$ with the maximal absolute value of reduced cost.

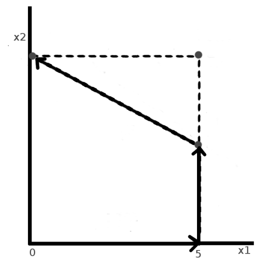
Theorem: Klee-Minty Cube (KMC) Under the previous assumption, the simplex algorithm can take "exponentially" (related with the input size) many steps.

Proof (construction/deformation of the cube): $\mathcal{P} = \{x \in \mathbb{R}^n \mid Ax \leq b\}$ is a deformation of the cube.

$$\begin{array}{ll} \max & 2x_1 + x_2 \\ \text{subject to} & x_1 \leq 5 \\ & 4x_1 + x_2 \leq 25 \\ & x_1, x_2 \geq 0 \end{array}$$

$\rightarrow$ by adding slack variables:

Figure 6: Cube degeneracy: skip one vertice. It can be seen as a "cube", but the skipped vertex is not part of the problem.



$$
\begin{aligned}
\max \quad & 2x_1 + x_2 \\
\text{subject to} \quad & x_1 + z_1 = 5 \\
& 4x_1 + x_2 + z_2 = 25 \\
& x, z \geq 0
\end{aligned}
$$

$\rightarrow \{z_1, z_2\}$ are the basis.

Pivot column $\{1\}$ (original vertex $(0,0)$. In the columnm $\{1\}$ the maximum quantity is $x_1 = 5 - z_1$. Change the value of x_1 on the objective function and constraints. We can ignore positive constants on the objective function, they are irrelevant for maximization.

$$
\begin{aligned}
\max \quad & x_2 - 2z_1 \\
\text{subject to} \quad & x_1 + z_1 = 5 \\
& -4z_1 + x_2 + z_2 = 5 \\
& x, z \geq 0
\end{aligned}
$$

$\rightarrow \{x_1, z_2\}$ are the basis, original vertex $(5,0)$, new pivot column is $\{2\}$. We choose by trying to eliminate positive costs.
$x_2 = 5 - z_2 + 4z_1$

$$
\begin{aligned}
\max \quad & -z_2 + 2z_1 \\
\text{subject to} \quad & x_1 + z_1 = 5 \\
& x_2 + z_2 - 4z_1 = 5 \\
& x, z \geq 0
\end{aligned}
$$

$\rightarrow \{x_1, x_2\}$ are the basis, original vertex $(5,5)$.
$z_1 = 5 - x_1$

$$\begin{aligned} \max \quad & -z_2 - 2x_1 \\ \text{subject to} \quad & x_1 + z_1 = 5 \\ & x_2 + z_2 + 4x_1 = 25 \\ & x, z \geq 0 \end{aligned}$$

$\rightarrow \{z_1, x_2\}$ are the basis, original solution $(0, 25)$. There is no positive costs in the objective function.

The generalization

Add z as slack variables and turn inequalities in equalities.

$$\begin{aligned} \max \quad & 2^{n-1}x_1 + \cdots + x_n \\ \text{subject to} \quad & z_1 + x_1 \leq (=)5 \\ & z_2 + 4x_1 + x_2 \leq (=)5 \times 5 \\ & \vdots + 8x_1 + 4x_2 + x_3 \leq (=)5 \times 5 \times 5 \\ & \vdots + \vdots \leq (=)\vdots \\ & z_n + 2^n x_1 + 2^{n-2}x_2 + \cdots + 2^2 x_{n-1} + x_n \leq (=)5^n \\ & x, z \geq 0 \end{aligned}$$

The feasible region of this polyhedron is called the Klee-Mint Cube (KMC).

Observations:

1. All basis of KMC are non-degenerate.
2. For any basis, either x_i or z_i is on the basis
3. The unique optimal solution is $x_1 = 0 = x_2 = \cdots = x_{n-1}$; $x_n = 5^n$ (maximum weight in the last variable).
4. Starting from zero and pivoting in a column with highest reduced cost coefficient require $2^n - 1$ many pivoting steps.

Finding an initial basis

Lemma: Let $\mathcal{P} = \{x \geq 0 \mid Ax = b\}$. By solving a linear optimization problem, we can find an initial bfs if it exists or prove that it does not exist.

Proof: First multiply every row by -1 if $b_i < 0$. Then, wlog $b \geq 0$. Consider auxiliar system: $\alpha = \min \sum_{i=1}^{m} y_i$ st $Ax + Iy = b$, $x \geq 0$, $y \geq 0$. This system has a bfs, because the initial basis is $x = 0$, $y = b$. A lower bound for α is zero, otherwise, there is no bfs. If $\alpha > 0$, then no solution exists. If $\alpha = 0$ the SA terminates with a basic solution only using x variables.

The Dual Simplex Algorithm

Consider $\max\{y^T b \mid A^T y \leq c\}$ as the dual. Let y^* be the dual bfs associated with some basis $B \subseteq \{1, \ldots, m\}$. Such bfs has the property: $y^* = c_B^T A_B^{-1}$, and dual feasible implies $\bar{c} \geq 0$.

Lemma: Suppose $\exists i \in B$ st $\bar{b_i} < 0$:

- We look into the matrix A, if $\bar{A_{ij}} \geq 0 \ \forall j \in N$ then $\max\limits_{y \in \mathcal{D}} y^T b = +\infty$
- Otherwise, $\lambda^* = \min\{\frac{-\bar{c_j}}{\bar{A_{ij}}} \mid \bar{A_{ij}} < 0,\ j \in N\}$ and $(y^*)^T = (y^*)^T - \lambda^*(A_B^{-1})_i$ is feasible for the dual. $b^T y = b^T y^* - \lambda^* \bar{b_j}$, i.e, $b^T y > b^T y^*$, where $\lambda^* = \frac{-\bar{c_k}}{\bar{A_{ik}}}$ (k is not in the basis).
  $B \setminus \{i\} \cup \{k\}$ is a new basis.

Proof: Let $y^T = (y^*)^T - \lambda(A_B^{-1})_i$ where $\lambda > 0$. $y^T A = (y^*)^T A - \underbrace{\lambda}_{\geq 0} \underbrace{\bar{A_i}}_{\geq 0} \leq c^T$. In the otherwise case: $y^T A = (y^*)^T A - \lambda^* A_i$. If $\bar{A_{ij}} \geq 0 \rightarrow y^T A \leq c_j$, otherwise, if $\bar{A_{ij}} < 0 \rightarrow y^T A \leq (y^*)^T A_{\cdot j} \frac{-\bar{c_j}}{-\bar{A_{ij}}} \bar{A_{ij}} = \underbrace{c_B^T A_B^{-1} A_{\cdot j}}_{y^*} + \underbrace{c_j - c_B^T A_B^{-1} A_{\cdot j}}_{\bar{c_j}} = c_j$. $B \setminus \{i\} \cup \{k\}$ being a basis is the same argument as in lecture 8, pivoting-theorem.

The major simplex steps in the dual setting

1. Choose a dual bfs in the dual setting: y^*
2. $\bar{b} = A_B^{-1} b$
3. If $\bar{b} \geq 0$, stop: optimal
4. Otherwise, let $i \in B$, $\bar{b_i} < 0$
   if $\bar{A_{ij}} \geq 0 \ \forall j \in N$, stop: unbounded
   otherwise, compute $\lambda^* = min\{\frac{\bar{c_j}}{-\bar{A_{ij}}} \mid \bar{A_{ij}} < 0 \ j \in N\}$ and an index $k \in N$, st $\lambda^* = \frac{-\bar{c_k}}{\bar{A_{ik}}}$.
5. Swap indices and update solutions, $B = B \setminus \{i\} \cup \{k\}$, $(y^*)^T = (y^*)^T - \lambda^*(A_B^{-1})_i$.

Simplex Algorithm IV

The revised Simplex Algorithm

Implementational issue: update algorithmic for basis inverses Mathematics is unchanged. Core: "elementary row operations".

Definition (Elementary Row Operation): Given a matrix $A \in \mathbb{R}^{m\times n}$, the operation of adding a multiple of $\lambda \neq 0$ of rows i to row k is an elementart row operation (ERO). When $i = k$, we are doing a scale operation. The resulting matrix $\bar{A} = Q \times A$, when $Q = I^{m\times m} + D$, where $D_{k,i} = \lambda$ and $D_{r,s} = 0$ elsewhere.

$$Q = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \lambda & 0 \\ \vdots & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

λ is on the intersection of row k and column $i \to D_{k,i} = \lambda$

Q is invertible

Remark: By applying a sequence of l elementary row operations to a matrix A, it gives us $\bar{A} = Q_l \times Q_{l-1} \times \dots \times Q_1 A$. Where each Q_i is of the form introduced before.

Observation: Let $Q = Q_l \times Q_{l-1} \times \dots \times Q_1$, then Q^{-1} exists. $Q^{-1} = Q_1^{-1} \times Q_2^{-1} \times \dots \times Q_l^{-1}$. This is true because $Q \times Q^{-1} = (Q_l \times \dots \times Q_1)(Q_1^{-1} \times \dots \times Q_l^{-1}) = I$.

Application to basis update: Let $\mathcal{P} = \{x \in \mathbb{R}^n \mid Ax \leq b, x \geq 0\}$. $A \in \mathbb{Q}^{m\times n}$ of full row rank. Change the basis takes in the worst case, m row operations.

Assume basis $B \subseteq \{1, \dots, m\}$ is given, i.e, $|B| = m$, A_B^{-1} exists. Suppose we apply a pivot operation. This leads us to a new basis $\bar{B}$, such that $l \in B$, $l \notin \bar{B}$, $j \in N$, $j \notin B$, $j \in \bar{B}$, $l \in N$.
Wlog $B = \{1, \dots, m\}$. $A_{\bar{B}} = [A_{\cdot 1}| \dots |A_{\cdot, l-1}|A_{\cdot j}|A_{\cdot, l+1}| \dots |A_{\cdot m}]$

$$A_B^{-1}A_{\bar{B}} = \left[\begin{array}{ccc|c|ccc} 1 & & 0 & & 0 & 0 & 0 \\ & \ddots & 1 & u & 1 & \ddots & \\ 0 & & & & 0 & & 1 \end{array}\right] = [e_1| \dots |e_{l-1}|u|e_{l+1}| \dots |e_m]$$

$u = A_B^{-1}A_{\cdot j}$

Where e_i is the i-th unit vector in $\mathbb{R}^m$.

Goal: apply a sequence of elementary rows operations to the matrix $A_B^{-1}A_{\bar{B}}$, so the new matrix becomes the identity I. For $i \neq l$: add $(\frac{-u_i}{u_l})$ times row l to row i.
Replace row l by $\frac{1}{u_l}$ times row l.

Observation: From our discussions there exists an invertible matrix Q, such that $QA_B^{-1}A_{\bar{B}} = I \to QA_B^{-1} = A_{\bar{B}}^{-1}$.

An iteration of the revised SA

1. Let B be a basis. Assume we know A_B^{-1}. $(x_B, x_N) = (A_B^{-1}b, 0)$.

2. Compute $p^T = c_B^T A_B^{-1}$. Compute $\bar{c_j} = c_j - p^T A_j$ (the reduced cost) $\forall j \in N$. If $\bar{c_j} \geq 0 \forall j$, stop: optimal.

3. Compute $u = A_B^{-1} A_{\cdot j}$, $\forall j \in N$, such that $\bar{c_j} < 0$. If $u_i \leq 0 \forall i$, stop: unbounded.

4. Compute $\lambda^* = \min\{\frac{x_i}{u_i} : i \in B,\, u_i > 0\}$.

5. Determine $l \in B$ such that $\lambda^* = \frac{x_l}{u_l}$. Determine $\bar{B} = B \setminus \{l\} \cup \{j\}$. $x_{\bar{B}} = \{x_i - \lambda^* u_i$, for all $i \in B \setminus \{l\}$, $x_j = \lambda^*\}$.

6. Let Q be the matrix from the previous observation. Then, $A_{\bar{B}}^{-1} = Q A_B^{-1}$.

Degeneracy and anticycling rules

(Primal) $\min\{c^T x \mid Ax = b, x \geq 0\}$.
B is a basis, degenerate, i.e, $A_B^{-1} b \not> 0$, there is some components that are zero.
What can happen, let $j \in N$, $\bar{c_j} > 0$. $\lambda^* = \max\{\lambda \in \mathbb{R}_+ \mid x^* + \lambda d \in \mathcal{P}\}$, where d is the j-th non-basis direction.
If some components are zero, λ^* can be zero, and we cannot change our bfs. However, we can still have another basis representation, this can lead to a loop (cycle).

Definition - degenerate pivot operation Let B be a basis. A pivot operation is called degenerate if the objective function value after the basis switch does not change.

Lemma: A pivot operation is degenerate $\iff$ $\lambda^* = 0$ $\iff$ the bfs x^* does not change.

Proof: $\lambda^* = 0 \to$ bfs does not change $x^* = x^* + \underbrace{\lambda^*}_{=0} d = x^*$.
Suppose objective function does not change after the pivot. Then $c^T(x^* + \lambda^* d) = c^T x^* \to \lambda^* \underbrace{c^T d}_{\bar{c_j}} = 0 \to \lambda^* \underbrace{\bar{c_j}}_{<0} = 0 \to \lambda^* = 0$.

Observation: For any deterministic pivot-rule (an algorithm that tells us which $j \in N$, $\bar{c_j} < 0$ to select and which pivot-row to select, in case it is not unique), either the SA terminates in finite time or it cycles through degenerate basis.

Proof: We have at most $\binom{m}{n}$ possible basis. Either the SA visits each basis at most once (finite termination of the algorithm) or one basis B is seen at least twice. Consider subsequence of basis $B \mapsto B_1 \mapsto \cdots \mapsto B \to c^T(x_B^*) \geq c^T(x_{B_i}^*) \geq \cdots \geq c^T(x_B^*)$, where x_B^* is the bfs in $\mathbb{R}^n$ associated with $B \to c^T(x_{B_i}^*) = c^T(x_B^*) \underbrace{\to}_{\text{lemma}}$ we cycle through degenerate basis.

We need anticycling rules

Definition - lexicographic order Let $u, v \in \mathbb{R}^n$, $u <_{lex} v$ if there exists $k \in \{1, \ldots, n\}$, such that $u_i = v_i$ $\forall i = 1, \ldots, k-1$.

Definition - lexicographic pivot rule Let B a basis and $T \in \mathbb{R}^{m+1 \times n+1}$ the tableu matrix.

1. Choose $j \in N$, $\bar{c_j} < 0$ arbitrarily.
2. Among all rows k such that $\frac{x_k^*}{|A_{kj}|} = \lambda^* = \min\{\frac{x_i^*}{|A_{ij}|}$, choose one such that the row vector $\frac{T_k}{|A_{kj}|}$ is minimal lexicographically.

Theorem: The SA with a lexicographic pivoting rule always terminates in finite time.

Proof: Start with basis and bfs. $B = \{1, \ldots, m\}$, then $\bar{A} = (I, \bar{A_N}) \rightarrow$ every row $\bar{A_{i\cdot}} >_{lex} 0 \rightarrow T_{i\cdot} >_{lex} 0$, $\forall i \geq 1$.

Claim: In the course of pivoting $T_{i\cdot} >_{lex} 0$, $\forall i \geq 1$

Proof: Pick $j \in N$, $\bar{c_j} < 0$. Pick $k \in B$ st $\frac{T_k}{\bar{A_{kj}}}$ lexicographically smallest. Start basis $\{1, \ldots, m\}$, wlog:
$T_{i\cdot} >_{lex} 0 \forall i \geq 1$.
Claim: (exercise) Invariant property $T_{i\cdot} >_{lex} 0 \forall i \geq 1$. We know $\bar{c_j} < 0$ and hence $T_{0\cdot} - \frac{\bar{c_j}}{\bar{A_{kj}}} \underbrace{T_{k\cdot}}_{\text{pivot row}} >_{lex} T_{0\cdot}$. So, no basis is visited twice.

Interior Point Methods

Preliminaries

Lemma: Let f be a function. $dom(f) \mapsto \mathbb{R}$ convex and continuously differentiable. Then, for $x^* \in dom(f)$: $f(x^*) = \min_{x \in dom(f)} f(x) \iff \nabla f(x^*) = 0$

Observation: In $\mathbb{Z}$ we do not have a criteria like this. Proof on lecture 13.

Lemma: Assume f strictly convex. Suppose Q is the domain, $Q \subseteq dom(f)$ is compact, (i.e, if you take a sequence, look to the limit, limit lies on the sequence). Q is convex. Then, there exists a unique $x^* \in Q$ such that $f(x^*) = \min_{x \in Q} f(x)$.

Proof: The existence of x^* follows from compactness of $Q \subseteq dom(f)$. For $x, y \in Q$, $x \neq y$, let $z = \frac{1}{2}x + \frac{1}{2}y$. Then, from the strictly convexity, $f(z) < \frac{1}{2}f(x) + \frac{1}{2}f(y) \leq max\{f(x), f(y)\}$. This implies that both x and y cannot be both minimizers (optimal solutions).

Logarithmic Barriers

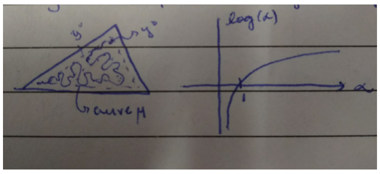
Consider primal-dual pair of linear optimization problems: Primal (P) $p^* = \max\{c^T x \mid Ax = b, x \geq 0\}$. Dual (D) $d^* = \min\{b^T y \mid A^T y \geq c, y \in \mathbb{R}^m\}$.

Assumptions:

- A has full row rank (i.e., looking at the rows of the linear system, either we can remove redundant constraints or the problem is feasible).
- $D = \{y \in \mathbb{R}^m \mid A^T y \geq c\}$ is bounded (test recession cone $A^T y \geq 0$, farka's lemma).
- There exists an "interior point" in D, i.e., $\exists y$ such that $\underbrace{A^T y > c}_{\text{component-wise}}$. i.e., the polyhedron is "full dimensional".

Definition: Consider the function $\phi : \mathbb{R}^m \mapsto \mathbb{R}$. $\phi(y) = -\sum_{j=1}^{n} log(y^T A_{\cdot j} - c_j)$ is called a log barrier for D. Log barrier only works for the dual.

The log barrier method prevents to reach the boundary, so, it prevents to reach a vertex.



- For every $\mu > 0$, let $\psi_\mu : \mathbb{R}^m \mapsto \mathbb{R}$, $\psi_\mu(y) = b^T y + \frac{1}{\mu}\phi(y)$
- The curve $\mu \mapsto y^*(\mu) \in \mathbb{R}^m$. $y^*(\mu)$ is a point that attains the minimum. $y^*(\mu) = \arg \min\limits_{y \in dom(\psi_y)} \psi_m u(y)$ is called the central path.

Lemma:

1. $dom(\psi_\mu) = dom(\phi) = \{y \in \mathbb{R}^m \mid y^T A > c\}$.
2. ψ_μ is strictly convex (Hessian matrix).
3. $y^*(\mu)$ is uniquely defined.
4. $y^*(\mu)$ is strictly feasible for (D).
5. $\nabla \psi_\mu(y^*(\mu)) = 0$.

Proof:

1. obvious (related with log function)

2. $\nabla\phi(y) = -\sum_{j=1}^{n}(\frac{1}{y^T A_{\cdot j} - c_j})A_{\cdot j}$.
$\nabla^2\phi(y) = \sum_{j=1}^{n}(\frac{1}{(y^T A_{\cdot j} - c_j)^2})(A_{\cdot j}A_{\cdot j}^T) \geq 0$, i.e, $\forall x \in \mathbb{R}^m \setminus \{0\}$, $x^T\nabla^2\phi(y)x \geq 0$: (PSD).
Assume $x^T\nabla^2\phi(y)x = 0 \iff 0 = A_{\cdot j}^T x \; \forall j \in \{1,\ldots,n\} \iff x \in ker(A^T)$, but this contradicts full row rank A, unless $x = 0 \rightarrow \forall y \in dom(\phi)$: $\nabla^2\phi(y) > 0 \rightarrow \phi$ is strictly convex $\rightarrow \psi_\mu$ is strictly convex.

3. Claim: the minimum $\psi_\mu(y)$ is attained in $\{y \in \mathbb{R}^m \mid y^T A \geq \bar{c_j}\}$ ($= \bar{D}$), where $\bar{c_j}$ are some values not the reduced cost.
(D) is bounded, $dom(\phi)$ is bounded, $\bar{D}$ is compact and convex. Then, the statement 3 follows from preliminaries lemma.
Proof of the claim: $\underbrace{b^T y \text{ is bounded over } dom(\psi_\mu)}_{\text{the domain itself is bounded}} \rightarrow \exists\beta_j \in \mathbb{R}$ st $log(y^T A_{\cdot j} - c_j) \leq \beta_j \; \forall y \in dom(\psi_\mu)$ and $y_0 \in dom(\psi_\mu)$ and $k \in \{1,\ldots,n\}$ we obtain $(*) b^T y - \sum_{j=\{1,\ldots,n\}\setminus k} \frac{1}{\mu}\beta_j - \frac{1}{\mu}log(\bar{c_k} - c_k) \underbrace{>}_{\exists\bar{c_k}\in\mathbb{R}} \psi_\mu(y_0)$. This implies $\forall y \in dom(\psi_\mu)\; \psi_\mu(y) \geq b^T y - \sum \frac{1}{\mu}\beta_j - \frac{1}{\mu}log(y^T A_{\cdot k} - c_k) \geq (*) > \psi_\mu(y_0)$.

4. clear

5. from lemma in preliminaries

Lemma: let $\mu > 0$, the point $x^*(\mu)$ defined via $x_j^*(\mu) = \frac{1}{\mu(y^*(\mu)^T A_{\cdot j} - c_j)}$ $\forall j$ satisfies:

1. $x^*(\mu)$ is (strictly) feasible for (P).

2. $gap(x^*(\mu), y^*(\mu)) = b^T y^*(\mu) - c^T x^*(\mu) = \frac{n}{\mu}$. We can choose μ so the gap between (P) and (D) is really small. The catch: must be possible to find y^*.

Proof:

1. $y^*(\mu)^T A_{\cdot j} - c_j > 0 \forall j \rightarrow x^*(\mu)j > 0$.
From previous lemma: $0 = \nabla\psi_\mu(y^*(\mu)) = b - \sum_{j=1}^{n} \frac{1 \times A_{\cdot j}}{\mu(y^*(\mu)^T A_{\cdot j} - c_j} = b - \sum_{j=1}^{n} A_{\cdot j} x_j^*(\mu) = b - Ax^*(\mu) \rightarrow b = Ax^*(\mu)$.

2. $y^*(\mu)^T b - c^T x^*(\mu) = y^*(\mu)^T A x^*(\mu) - c^T x^* \mu = (y^*(\mu)^T A - c^T)x^*(\mu) = \sum_{j=1}^{n}(y^*(\mu)^T A_{\cdot j} - c_j)x_j^*(\mu) \underbrace{=}_{\text{from } x^*(\mu) \text{ formula}} \sum_{j=1}^{n}(\frac{1}{\mu}) = \frac{n}{\mu}$.

Our analysis allow us to formulate the Interior Point Algorithm:

1. Choose $\mu_0 > 0$, $k > 1$

2. Solve $y^*(\mu_0) = \arg \min_{y \in dom(\psi_{\mu_0})} \psi_{\mu_0}(y)$.

3. For $t \geq 0$ do:
   if $\frac{n}{\mu_t} \leq \varepsilon$: quit
   $\mu_{t+1} = k\mu_t$
   solve $y^*(\mu_{t+1}) = \arg \min_{y \in dom(\psi_{\mu_{t+1}})} \psi_{\mu_{t+1}}(y)$

Lemma: Let $\varepsilon > 0$. Suppose we can solve step 2 and 3 (third line) with some oracle, then Integer Point Algorithm requires at most $T = \lceil \frac{ln(\frac{n}{\mu_0 \varepsilon})}{ln(k)} \rceil = \lceil log_k(\frac{n}{\mu_0 \varepsilon}) \rceil$ iterations to find a pair of primal/dual feasible solutions whose gap is at most ε.

Proof: After T iterations the gap is $\frac{n}{\mu_t} = \frac{n}{k^T \mu_0} = \frac{n}{k^{\lceil log_k(\frac{n}{\mu_0 \varepsilon}) \rceil} \mu_0} \leq \frac{n}{k^{log_k(\frac{n}{\mu_0 \varepsilon})\mu_0}} = \frac{n}{\frac{n}{\mu_0 \varepsilon}\mu_0} = \varepsilon$.

Convex Optimization and the Newton Method

f is always a convex function from now on.

Convex optimization

Basic properties $f* = \min_{x \in Q} f(x)$, f convex function, Q convex set, $Q \subseteq dom(f)$, f^* is the optimal solution.

Remember: to identify optimal solutions, we need to evaluate conditions (in any class of optimization problems).

Lemma: f continuously differentiable on $dom(f)$. $x^* \in Q$ is a minimizer $\iff$ $\nabla f(x^*)^T(y - x^*) \geq 0$, $\forall y \in Q$.

Proof:

- "$\leftarrow$":
  $x^* \in Q$ satisfies $\nabla f(x^*)^T(y - x^*) \geq 0$, $\forall y \in Q$, then $\underbrace{f(y) \geq f(x^*) + \underbrace{\nabla f(x^*)^T(y - x^*)}_{\geq 0} \geq}_{f \text{ convex and diff'ble}}$ $f(x^*)$, $\forall y \in Q$, thus x^* is a minimizer.

- "$\rightarrow$":
  Assume $f(x^*) = f^*$ and suppose $\exists y \in Q$ st $\nabla f(x^*)^T(y - x^*) < 0$.
  $\phi : \mathbb{R} \mapsto \mathbb{R}$, $\phi(\lambda) = f(\lambda y + (1 - \lambda)x^*)$. ϕ is convex.
  $\phi'(\lambda) = \nabla f(\lambda y + (1 - \lambda)x^*)^T(y - x^*)$.

$\phi'(0) = \nabla f(x^*)^T(y - x^*) \underbrace{< 0}_{\text{by assumption}}$.

$\exists\lambda > 0$ such that $\phi(\lambda) < \phi(0)$, this is a contradiction.

Lemma: f continuously differentiable on $dom(f)$. $x^* \in dom(f)$ attains $f^* \iff \nabla f(x^*) = 0$.

Proof:

- "$\leftarrow$":

  if $x^* \in dom(f)$ and $\nabla f(x^*) = 0$, then, $\forall y \in dom(f)$, $f(y) \geq f(x^*) + \nabla f(x^*)^T(y - x^*) = f(x^*)$.

- "$\rightarrow$":

  Suppose $x^* \in dom(f)$ attains f^*. Choose some positive number $t > 0$ such that $y = x^* - t\nabla f(x^*) \in dom(f)$.
  From previous lemma, $0 \leq \nabla f(x^*)^T(y - x^*) = -t\underbrace{||\nabla f(x^*)||^2}_{\to 0} \leq 0$.

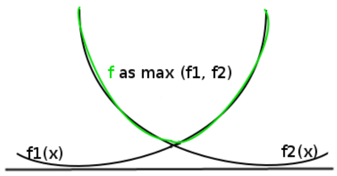


Figure 7: f now is convex, we cannot replace $\max$ with $\min$ because it will result in a concave function instead.

Lemma: Let $f_1, \ldots, f_m$: be convex functions. Let $D = \bigcap_{i=1}^m dom(f_i) \subseteq \mathbb{R}^n$. $f(x) = \max\{f_1(x), \ldots, f_m(x)\}$ $\forall x \in D$ is convex.

Proof: $\forall x, y \in D$: $f(\lambda x + (1-\lambda)y) = \max\{f_i(\lambda x + (1-\lambda)y)\colon i = 1, \ldots, m\} \leq \max\{\lambda f_i(x) + (1-\lambda)f_i(y)\colon i = 1, \ldots, m\} \leq \max\{\lambda f_i(x)\colon i = 1, \ldots, m\} + \max\{(1-\lambda)f_i(y)\colon i = 1, \ldots, m\} = \lambda\max\{f_i(x)\colon i = 1, \ldots, m\} + (1-\lambda)\max\{f_i(y)\colon i = 1, \ldots, m\} = \lambda f(x) + (1-\lambda)f(y)$

Definition: strongly convexity (not strictly convexity) $f : dom(f) \mapsto \mathbb{R}$ is strongly convex with modulus $\sigma > 0$, if $\forall\lambda \in (0, 1)$ and $x, y \in dom(f)$: $f(\lambda x + (1-\lambda)y) + \frac{\sigma}{2}\lambda(1-\lambda)||y - x||^2 \leq \lambda f(x) + (1-\lambda)f(y)$.

Remark:

- Let f be continuously differentiable on $dom(f) \neq \emptyset$ convex. Then f is strongly convex with modulus $\sigma \iff f(y) \geq f(x) + \nabla f(x)^T(y - x) + \frac{\sigma}{2}||y - x||^2$ $\forall x, y \in dom(f)$.

- If f is strongly convex with modulus $\sigma \iff \nabla^2 f(x) \geq \sigma I$, i.e., $\nabla^2 f(x) - \sigma I \geq 0$ (PSD).

Example: $f(x) = ||x||^2$ is strongly convex with modulus 2 (diagonal of Hessian Matrix is 2).

Remember: Hessian Matrix is a square matrix of second order partial derivatives.
$H_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$. Hessian matrix of a convex function is PSD. If Hessian is positive definite at x, then f attains an isolated local minimum at x. If it is negative, then f attains an isolated local maximum. If it has both positive and negative eigenvalues, then x is a saddle point for f.

Lemma: if f is strongly convex and twice differentiable with modulus σ, then for all $x \in dom(f)$, $\nabla^2 f(x)^{-1}$ (inverse of the Hessian) exists and $||\nabla^2 f(x)^{-1} y||^2 \leq \frac{||y||^2}{\sigma^2} \forall y \in dom(f)$.

Proof: Let $A = \nabla^2 f(x)$ be the Hessian Matrix. $A \geq \sigma I$, A is symmetric and A^{-1} exists.
$\forall y \in dom(f)$: $y^T A y \geq \sigma ||y||^2$ (from PSD fact)
$y^T A^{-1} y = y^T A^{-1} A A^{-1} y \geq \sigma ||A^{-1} y||^2$ (*)
$||y||^2 = y^T y = y^T A^{-\frac{1}{2}} A^{\frac{1}{2}} A^{\frac{1}{2}} A^{-\frac{1}{2}} y = y^T A^{-\frac{1}{2}} A A^{-\frac{1}{2}} y \geq \sigma ||A^{-\frac{1}{2}} y||^2$
$= \sigma y^T A^{-1} y \underbrace{\geq}_{*} \sigma ||A^{-1} y||^2$.

The Newton Method

$f^* = \min_{x \in dom(f)} f(x)$
"no constraints" (convex optimization without constraints)
Assumptions:

- f is twice differentiable on $dom(f)$
- f is strongly convex with modulus σ
- There exists a constante $L > 0$ such that $||\nabla^2 f(x) - \nabla^2 f(y)||_\nabla \leq L||x - y||_2 \ \forall x, y \in dom(f)$, where $||A||_\nabla = sup\{\frac{||Az||_2}{||z||_2} : z \neq 0\}$
- $\exists x_0 \in dom(f)$ such that $||\nabla f(x_0)||_2 \leq \frac{\sigma^2}{L}$

Goal: find $x^* \in dom(f)$ such that $\nabla f(x^*) = 0$. (There are many situation where this is not easy to find.)
Develop an iterative scheme:

Approximate functions: $x \mapsto \nabla f(x_t) + \nabla^2 f(x_t)(x - x_t)$
Now, we set this function to be zero:
$\nabla f(x_t) + \nabla^2 f(x_t)(x - x_t) = 0 \iff x_{t+1} = x_t + \delta t$.
$\delta t = -[\nabla^2 f(x_t)]^{-1} \nabla f(x_t)$

Newton Method's Algorithm

1. Begin with x_0
2. For $0 \leq t \leq T$:
   compute $\delta t = -[\nabla^2 f(x_t)]^{-1} \nabla f(x_t)$ and set $x_{t+1} = x_t + \delta t$
3. return x_t

Theorem: $f(x_t) - f^* \leq \frac{2\sigma^3}{L^2}(\frac{1}{2})^{2^{T+1}}$**, i.e, under the above assumptions, the Newton method converges quadratically to the optimal value** f^***, provided that we start close enough (**$||\nabla f(x_0)||_2 \leq \frac{\sigma^2}{L}$**)**

Proof: We know $f(y) \geq \underbrace{f(x_t) + \nabla f(x_t)^T (y - x_t) + \frac{\sigma}{2}||y - x_t||^2}_{=g(y)}$ $\forall y \in dom(f)$.
$g(y)$ is strictly convex and attains its minimum at $y^* = x_t - \frac{1}{\sigma}\nabla f(x_t)$.
$\rightarrow f(y) \geq g(y) \geq g(y^*) = f(x_t) - \frac{1}{2\sigma}||\nabla f(x_t)||^2$
$\rightarrow f^* \geq f(x_t) - \frac{1}{2\sigma}||\nabla f(x_t)||^2 \rightarrow f(x_t) - f^* \leq \frac{1}{2\sigma}||\nabla f(x_t)||^2$

$$
\begin{aligned}
&||\nabla f(x_{t+1})|| = ||\nabla f(x_t + \delta t)|| \\
&= ||\nabla f(x_t + \delta t) - \nabla f(x_t) - \nabla^2 f(x_t)\delta t|| \\
&= || \textstyle\int_0^1 \nabla^2 f(x_t + s\delta t)\delta t ds - \nabla^2 f(x_t)\delta t|| \\
&= \textstyle\int_0^1 ||[\nabla^2 f(x_t + s\delta t) - \nabla^2 f(x_t)]\delta t ds|| \\
&\leq \textstyle\int_0^1 ||[\nabla^2 f(x_t + s\delta t) - \nabla^2 f(x_t)]\delta t|| ds \\
&\leq \textstyle\int_0^1 ||\nabla^2 f(x_t + s\delta t) - \nabla^2 f(x_t)||_{\nabla} \times ||\delta t|| ds \\
&\leq \textstyle\int_0^1 L \times \underbrace{||x_t + s\delta t - x_t||}_{\text{from nabla norm}} ||\delta t|| ds \\
&= \textstyle\int_0^1 Ls||\delta t||^2 ds \\
&= \frac{L}{2}||\delta t||^2 = \frac{L}{2}||\underbrace{\nabla^2 f(x_t)^{-1}\nabla f(x_t)}_{\text{definition of } \delta t}||^2 \\
&\underbrace{\leq}_{\text{lemma 3(?)}} \frac{L}{2\sigma^3}||\nabla f(x_t)||^2
\end{aligned}
$$

$\rightarrow f(x_t) - f^* \leq \frac{1}{2\sigma}||\nabla f(x_t)||^2 \leq \frac{2\sigma^3}{L^2}(\frac{1}{2})^{2^{T+1}}$.

Convex Optimization Duality I

The Lagrange function

Definition - Lagrange function Let f and $g_1, \dots, g_m$ continuous and convex functions. Let the vector $g(x) = \begin{pmatrix} g_1(x) \\ \vdots \\ g_m(x) \end{pmatrix} \in \mathbb{R}^m$. Let $D = \bigcap_{i=1}^m dom(g_i)$, $D \subseteq dom(f)$.

Consider

$$f^* = \min\{f(x) : g(x) \leq b\}$$

$b \in \mathbb{R}^m$ given. The function $Z : \mathbb{R}^m_+ \mapsto \mathbb{R}$: $Z(\lambda) = \min\limits_{x \in D}\{f(x) + \lambda^T(g(x) - b)\}$ is called the lagrange function for 14.3 (the previous minimization problem).

We want to maximize $Z(\lambda)$ to achieve f^*.

Lemma:

1. $Z(\lambda) \leq f^* \; \forall \lambda \geq 0$ (λ is a vector)
2. $Z : \mathbb{R}^m_+ \mapsto \mathbb{R}$ is concave

Proof:

1. Let x^* be optimal for 14.3, then $g_i(x^*) \leq b_i \forall i$
   $Z(\lambda) \leq \underbrace{f(x^*)}_{f^*} + \underbrace{\lambda^T}_{\geq 0} \underbrace{(g(x^*) - b)}_{\leq 0} \leq f^*$
2. Let $\lambda_1, \lambda_2 \in \mathbb{R}^m$ and $\alpha \in (0,1)$. $\lambda = \alpha\lambda_1 + (1-\alpha)\lambda_2$
   Let $\hat{x} \in D$ be such that $Z(\lambda) = f(\hat{x}) + \lambda^T(g(\hat{x}) - b)$
   In particular, $\forall i \in 1,2$: $Z(\lambda_i) \leq f(\hat{x}) + \lambda_i^T(g(\hat{x}) - b) \to \alpha Z(\lambda_1) + (1-\alpha)Z(\lambda_2) \leq \alpha(f(\hat{x}) + \lambda_1^T(g(\hat{x}) - b) + (1-\alpha)(f(\hat{x}) + \lambda_2^T(g(\hat{x}) - b)) = f(\hat{x}) + \underbrace{(\alpha\lambda_1 + (1-\alpha)\lambda_2)^T}_{\lambda^T}(g(\hat{x}) - b) = Z(\lambda)$.

How to determine $\max\limits_{\lambda \geq 0} Z(\lambda)$? Remember: $Z(\lambda)$ is concave!

Definition - subgradient Let f be continuously and convex. Given $x \in dom(f)$, a vector $s \in \mathbb{R}^n$ is called a *subgradient* if $f(y) - f(x) \geq s^T(y - x) \; \forall y \in dom(f)$, i.e., $\delta f(x) = \{s \in \mathbb{R}^n \mid s \text{ subgradient of } x\}$.

add a figure here - check lecture annotations

Lemma: Let f be convex and continuous. x^* is minimum of $f \iff \vec{0} \in \partial f(x^*)$.

Proof: $f(x) \geq f(x^*) \forall x \in dom(f) \iff f(x) - f(x^*) \geq 0 = 0^T(x - x^*)$.

Remark: existence of subgradients Let f be continuous and convex. Then $\forall x_0 \in dom(f)$, $\partial f(x_0) \neq 0$ and $\forall d \in \mathbb{R}^m$ we have that

$$\lim_{\varepsilon \to 0} \frac{f(x_0 + \varepsilon d) - f(x_0)}{\varepsilon} = \sup\{h^T d \mid h \in \partial f(x_0)\}$$

The projection (sub) gradient algorithm

Not polytime: pseudo polytime algorithm.
Easy version: you can use the gradient, but you can use (plug) any subgradient.

Motivation: compute $\max_{\lambda \geq 0} Z(\lambda)$ or directly 14.3. We want to show that both attains the same results, under some assumptions.
$f* = \min\{f(x) \mid x \in Q\}$ and x^* is a minimizer.

- Q is convex, non-empty, compact
- $Q \subseteq dom(f)$
- Q has diameter $D > 0$, i.e, $||x - y||_2 \leq D \; \forall x, y \in Q$

Projection oracle: input $y \in \mathbb{R}^n$, output $PO(y) \in Q$ attaining $\min\{\underbrace{||x - y||_2}_{\text{strictly convex, projection is unique}} \mid x \in Q\}$.
We know $||x - PO(y)|| \leq ||x - y|| \forall x \in Q$.

f is convex, Lipschitz-continuous and differentiable on $dom(f)$:

$\vec{0} \leq |f(x) - f(y)| \leq L||x - y|| \; \forall x, y \in dom(f) \to ||\nabla f(x)|| \leq L \; \forall x \in dom(f)$.

$\vec{0} = |\nabla f(x)^T (y - x)| \to y = \lambda \nabla f(x) + r, \; r \perp \nabla f(x)$.

Proof: proof as homework
tips: $f(x) - f(x_t) \geq \nabla f(x_t)^T (x - x_t)$. play with: $x = \lambda \nabla f(x_t) + r^{\perp \nabla f(x)}$.

Algorithm:

1. Choose $\varepsilon > 0$, $x_0 \in Q$. Set $\lambda = \frac{2\varepsilon}{5L}$, $T = \lceil (\frac{5LD}{4\varepsilon})^2 \rceil$. (Since it depends on the diameter, it is not polytime.)
2. For $t = 0, \ldots, T - 1$:

   compute $y_{t+1} = x_t - \underbrace{\frac{\lambda}{||\nabla f(x_t)||}}_{\text{scale, walks in the negative direction}} \nabla f(x_t)$, $x_{t+1} = PO(y_{t+1})$.
3. return best point $\bar{x} = \min\{f(x_t) \forall t\}$ that was generated.

Theorem: $f(\bar{x}) - f^* \leq \varepsilon$

Proof: $\|x_{t+1} - x^*\|^2 = \|PO(y_{t+1}) - x^*\|^2 \leq \|y_{t+1} - x^*\|^2 = \|x_t - \frac{\lambda}{\|\nabla f(x_t)\|}\nabla f(x_t) - x^*\|^2 = \|x_t - x^*\|^2 - 2\frac{\lambda \nabla f(x_t)^T}{\|\nabla f(x_t)\|}(x_t - x^*) + \lambda^2 = \|x_t - x^*\|^2 + \frac{2\lambda \nabla f(x_t)}{\|\nabla f(x_t)\|}(x^* - x_t) + \lambda^2 \leq \|x_t - x^*\| + \frac{2\lambda}{\|\nabla f(x_t)\|}(f(x^*) - f(x_t)) + \lambda^2 \leq \|x_t - x^*\|^2 + \frac{2\lambda}{L}(f^* - f(x_t)) + \lambda^2 \to f(x_t) - f^* \leq \frac{L}{2\lambda}(\|x_t - x^*\|^2 - \|x_{t+1} - x^*\|^2) + \frac{L\lambda}{2}$ $f(\bar{x}) - f^* \leq f(x_t) - f^*$ [recursive part?] $\to T(f(\bar{x}) - f^*) \leq \frac{L}{2\lambda}(\|x_0 - x^*\|^2 - \underbrace{\|x_T - x^*\|^2}_{\geq 0 \to \text{dropped}}) + T\frac{L\lambda}{2}$

$\leq \frac{L}{2\lambda}D^2 + T\frac{L\lambda}{2} \to f(x) - f^* \leq \underbrace{\frac{L}{2\lambda T}}_{=\frac{4\varepsilon}{5}}D^2 + \underbrace{\frac{L\lambda}{2}}_{=\frac{\varepsilon}{5}} = \varepsilon.$

Algorithm extends to subgradients The projection gradient method naturally extends to using subgradients s such that $||s||_2 \leq L$ instead of gradients.

algorithm as homework

Some further tool

Theorem: Let $f_1, f_2, \ldots, f_m$ be convex and twice continuous differentiable functions. Let $D = \cap_{i=1}^m dom(f_i) \neq \emptyset$. Let $f(x) = \max\{f_1(x), \ldots, f_m(x)\}$. (subfunctions are differentiable, but in some points of $f(x)$ we need subgradients.)

For $x \in D, I(x) = \{i \in \{1, \ldots, m\}, f(x) = f_i(x)\}$. *Then,* $\partial f(x) = conv(\nabla f_i(x) \mid i \in I(x))$.

Proof:

- "$conv(\nabla f_i(x) \mid i \in I(x)) \subseteq \partial f(x)$"

  Take $x_0 \in D$. $\forall i \in I(x_0) \to f_i(x) - f_i(x_0) \geq \nabla f_i^T(x - x_0) \forall x \in D$.
  $f(x) - f(x_0) \underbrace{\geq}_{i \in I(x_0)} f_i(x) - f_i(x_0)$ ($f(x_0) = f_i(x_0)$, but $f(x) \geq f_i(x)$)
  $\geq \nabla f_i(x_0)^T(x - x_0) \forall x \in D$
  $\to \nabla f_i(x_0) \in \partial f(x_0)$ $\forall i \in I(x_0) \to conv(\nabla f_i(x_0) \mid i \in I(x_0)) \subseteq \partial f(x_0)$.

- "$\partial f(x) \subseteq conv(\nabla f_i(x) \mid i \in I(x))$"

  Let $x_0 \in D$. Conversely, suppose $\exists s \in \mathbb{R}^n$, $s \in \partial f(x)$, $s \notin conv(\nabla f_i(x) \mid i \in I(x))$.
  From the separate hyperplane theorem, $\exists d \in \mathbb{R}^n$, $p \in \mathbb{R}$ such that $d^T h \leq p$, $\forall h \in conv(\nabla f_i(x) \mid i \in I(x))$,

  $$d^T s > p$$

  $f_i(x_0) < f(x_0) \forall i \notin I(x_0)$. $\exists \alpha > 0$ such that $f(x_0 + \varepsilon d) = f_k(x_0 + \varepsilon d)$ for some $k \in I(x_0)$ and $\forall 0 \leq \varepsilon \leq \alpha$.

  1. $f(x) - f(x_0) \geq s^T(x - x_0) \underbrace{=}_{x = x_0 + \varepsilon d} \varepsilon d^T s$

2. $f(x) - f(x_0) = f_k(x_0 + \varepsilon d) - f_k(x_0) = \nabla f_k(x_0)^T(\varepsilon d) + r(\varepsilon)$, where $r(\varepsilon)$ is a taylor approximation and quadratic, i.e, $r(\varepsilon) = \sigma(\varepsilon^2) \to r(\varepsilon) \leq \beta\varepsilon^2$ for some constant β.

We know from 14.3: $\delta = d^T s - d^T \nabla f_k(x_0) > 0$.
From (1) and (2): $\varepsilon d^T s \leq \varepsilon d^T \nabla f_k(x_0) + r(\varepsilon) \to \varepsilon\delta \leq r(\varepsilon) \leq \beta\varepsilon^2$, which is a contradiction when $\varepsilon \to 0$.

Convex Optimization Duality II

KKT theorem - optimal certificate

Theorem: Let $f, g_1, \dots, g_m$ be convex and twice differentiable functions. Let $dom(f) \supseteq \cap_{i=1}^m dom(g_i)$.

Slater condition: $\exists\bar{x}$ such that $g(\bar{x}) < b$ (for all components). Let $Q = \min\{f(x) = g(x) \leq b\}$. $x^* \in Q$ minimizes $f \iff \exists\lambda_1^*, \dots, \lambda_m^* \geq 0$ such that:

- $\lambda_i^* = 0 \forall i \notin I(x^*) = \{j : g_i(x^*) = b_i\}$
- $\nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla g_i(x^*) = 0$

Proof

- "$\to$"

  Let $f* = min\{f(x) \mid g(x) \leq b\}$ (P). Consider $\phi(x) = \max\{f(x) - f^*, g_1(x) - b_1, \dots, g_m(x) - b_m\}$. x^* solves (P) $\leftrightarrow$ x^* minimizes ϕ, and attains value $\phi(x^*) = 0$ (can't be less than zero because of $\phi(x)$.
  $\partial\phi(x^*) = conv(\nabla f(x^*), \nabla g_i(x^*),\ i \in I(x^*) \leftrightarrow \exists\alpha_0 \geq 0$ and $\alpha_i \geq 0 \forall i \in I(x^*)$, such that $\alpha_0 + \sum_{i \in I(x^*)} \alpha_i = 1$ and $\alpha_0 \nabla f(x^*) + \sum_{i \in I(x^*)} \alpha_i \nabla g_i(x^*) = 0$ **(**)**
  Suppose $\alpha_0 = 0 \forall i \in I(x^*)$. We know $\underbrace{g_i(x) - g_i(x^*)}_{=g_i(x)-b_i} \geq \nabla g_i(x^*)^T(x - x^*)$

  $0 = 0^T(x - x^*) \underbrace{=}_{**} \sum_{i \in I(x^*)} \alpha_i \nabla g_i(x^*)^T(x - x^*) \leq \sum_{i \in I(x^*)} \alpha_i [\underbrace{g_i(x) - b_i}_{\leq 0}] \forall x$, and this is a contradiction. This is impossible for $x = \bar{x}$: slater point, so $\alpha_0 \neq 0 \to \alpha_0 > 0$.
  Define now $\lambda_i^* = 0 \forall i \notin I(x^*)$, $\lambda_i^* = \frac{\alpha_i}{\alpha_0} \forall i \in I(x^*)$. $0 = \alpha_0 \nabla f(x^*) + \sum_{i \in I(x^*)} \alpha_i \nabla g_i(x^*) \underbrace{=}_{\text{by dividing by } \alpha_0} \nabla f(x^*)$ $\sum_{i \in I(x^*)} \frac{\alpha_i}{\alpha_0} \nabla g_i(x^*) = \nabla f(x^*) + \sum_{i \in I(x^*)} \lambda_i^* \nabla g_i(x^*)$ and $\lambda_i^* = 0 \forall i \notin I(x^*)$.

- "$\leftarrow$"

Remark: The KKT point is x^* such that $\nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla g_i(x^*)$ minimizes $f + \sum_{i=1}^m \lambda_i^*(g_i(x) - b_i)$ over D.

Duality for convex optimization problems

Theorem Convex Optimization Duality Under assumption/notation of KKT, $f^* = min\{f(x) \mid g(x) \leq b\} = \max_{u \geq 0, u_0 \in \mathbb{R}} \{-u^T b + u_0 \mid -u^T g(x) + u_0 \leq f(x), \quad \underbrace{\forall x \in D}_{\text{infinitely many constraints}} \quad \}$.

Proof:

$$\max_{u \geq 0, u_0 \in \mathbb{R}} \{-u^T b + u_0 \mid -u^T g(x) + u_0 \leq f(x) \forall x \in D\}$$
$$= \max_{u \geq 0, u_0 \in \mathbb{R}} \{-u^T b + u_0 \mid u_0 \leq f(x) + u^T g(x) \forall x \in D\}$$

u_0 is indeed a lower bound

$$= \max_{u \geq 0, u_0 \in \mathbb{R}} \{-u^T b + u_0 \mid u_0 \leq \min_{x \in D} \{f(x) + u^T g(x)\}\}$$

we want to make u_0 as big as possible and then remove any gap so $u_0 = min\{f(x) + u^T g(x)\}$

$$= \max_{u \geq 0} \{-u^T b + \min_{x \in D} \{f(x) + u^T g(x)\}\}$$
$$= \max_{u \geq 0} \{\min_{x \in D} \{f(x) + u^T [g(x) - b]\}\} (*)$$

From lecture #14: $Z(\lambda) = \min_{x \in D} \{f(x) + \lambda^T [g(x) - b]\}$, $\lambda \geq 0, Z(\lambda) \leq f^*$ and Z concave in λ. Then, $(*) = \max_{\lambda \geq 0} Z(\lambda) \leq f^*$, from the remark after KKT theorem, take KKT multipliers λ^* then, $Z(\lambda^*) = f^*$.

Remark: under slater condition, theorem of convex optimization duality allow us to rediscover LP-duality: $\min_{Ax \geq b} c^T x = \max_{stA^T y = c, y \geq 0} b^T y$.

Proof: $f(x) = c^T x, g(x) = b - Ax$

$\min\{f(x) \mid g(x) \leq 0\} (P) = \max_{u \geq 0, u_0 \in \mathbb{R}} \{u_0 \mid -u^T (b - Ax) + u_0 \leq c^T x \ \forall x\}$

$= \max_{u_0 \in \mathbb{R}, u \geq 0} \{u_0 \mid -u^T b + u_0 \leq [c^T - u^T A] x \ \forall x\}$

- Suppose does not exist $u \geq 0, c^T = u^T A \to \max_{u \geq 0, u_0 \in \mathbb{R}} \{u_0 \mid \dots\} = -\inf$
- Suppose $\exists u \geq 0, c^T = u^T A$. Then, $= \max_{u_0 \in \mathbb{R}, u \geq 0, u^T A = c^T} \{u_0 \mid -u^T b + u_0 \leq 0\} = \max\{u^T b \mid u \geq 0, u^T A = c^T\}$

Total Unimodular Matrices

Definition - Polyhedron integral Let $P \subseteq \mathbb{R}^n$ be a polyhedron. P is integral if every "minimal wrt inclusion" face of P contains integer points.

P is a polyhedron and also a polytope. P polytope $\to P = conv(v^1, \dots, v^t)$.
P is integral if the minimal faces (vertices) are integers, i.e, $v^i \in \mathbb{Z}^n$

Fact from linear algebra - Cramer's rule $A \in \mathbb{Q}^{n\times n}$, regular, $b \in \mathbb{Q}^n$. $Ax = b \iff x = A^{-1}b \iff \forall i = 1,\dots,n,\ x_i = \frac{det(A^i)}{det(A)}$, where $A^i = A^i_{\cdot l}\ \forall l \in \{1,\dots,n\} \setminus \{i\}$, $A^i_{\cdot i} = b$, i.e, A^i is the matrix formed by replacing the i-th column of A by the column vector b.

Definition - unimodular $A \in \mathbb{Z}^{m\times n}$ of full row rank is unimodular if the determinant of every basis of A ($m \times n$ regular submatrix) is equal to ± 1. A matrix $A \in \mathbb{Z}^{m\times m}$ of full row rank is unimodular if $det(A) = \pm 1$.

Definition - totally unimodular $A \in \mathbb{Z}^{m\times n}$ is called totally unimodular (TU) if every square submatrix of A has determinant 0, $+1$, or -1.

Examples:

- $\begin{bmatrix} 3 & 5 \\ 1 & 2 \end{bmatrix}$ is unimodular but not totally unimodular
- A is TU $\rightarrow A \in \{-1, 0, +1\}^{m\times n}$
- $A \in \{0, 1\}^{2\times w}$ is always TU: $\begin{bmatrix} 0 & 1 & 0 & 1 & \dots \\ 1 & 1 & 0 & 0 & \dots \end{bmatrix}$
- $A \in \{-1, 0, 1\}^{2\times n}$ is TU $\leftrightarrow$ it does not contain the submatrix $\begin{bmatrix} \pm 1 & \pm 1 \\ \pm 1 & -1 \end{bmatrix}$
- $A \in \{0, 1\}^{3\times n}$ is not always TU, for instance, determinant of $\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$ is 2.

Propositions

1. A is TU $\leftrightarrow [AI]$ is unimodular
2. A is TU $\leftrightarrow \begin{bmatrix} A \\ -A \\ I \\ -I \end{bmatrix}$ is TU
3. A is TU $\leftrightarrow A^T$ is TU

Proofs:

1. add proof
2. add proof
3. if A is TU, then every square submatrix of A has determinant $0, \pm 1$. Since $det(A) = det(A^T)$, every square submatrix of A^T has also determinant $0, \pm 1$.

Theorem - Unimodular Let $A \in \mathbb{Z}^{m\times n}$ of full row rank. A is unimodular if and only if the family of polyhedra $P(b) = \{x \in \mathbb{R}^n_+ \mid Ax = b\}$ is integral $\forall b \in \mathbb{Z}^m$ such that $P(b) \neq \emptyset$.

Proof

- direction "$\rightarrow$", Assume A is unimodular $\rightarrow$ let $b \in \mathbb{Z}^m$, $P(b) \neq \emptyset$. Take an extreme point $x^* \in P(b)$. $x^* = (x^*_B, x^*_N)$ for basis B, $x^*_B = A_B^{-1}b$, $x^*_N = 0$.
  $det[A_B] = \pm 1$ (because A is unimodular). From Cramer's rule $A_B\dot{x}^*_B = b$, x^*_B is integral. Hence, x has integer entries. Therefore, all extreme points of $P(b)$ are integral, i.e., $P(b)$ is integral.
- direction "$\leftarrow$", Assume x^* is integral. Suppose $P(b)$ is integral, $\forall b \in \mathbb{Z}^m$ such that $P(b) \neq \emptyset$. Let $B \subseteq \{1, \ldots, n\}$ be a basis. We want to show $det(A_B) \in \{-1, +1\}$. Let $b = A_B z + \underbrace{e_i}_{i\text{-th unit vector}}$,
  where z is integer ($z \in \mathbb{Z}^m$) such that $z + A_B^{-1}e_i \geq 0$. Consider $P(b)$ and the extreme point (x^*_B, x^*_N), $x^*_B = A_B^{-1}b = z + A_B^{-1}e_i$ which by assumption is feasible for $P(b)$ by $z + A_B^{-1}e_i \geq 0$. $z + A_B^{-1}e_i \in \mathbb{Z}^m \rightarrow A_B^{-1}e_i \in \mathbb{Z}^m \rightarrow A_B^{-1}$ is integral $\rightarrow det(A_B) \in \{-1, +1\}$.

Theorem - Totally Unimodular Let $A \in \mathbb{Z}^{m\times n}$. A is TU if and only if the family of polyhedra $P(b) = \{x \in \mathbb{R}^n_+ \mid Ax \leq b\}$ is integral $\forall b \in \mathbb{Z}^m$ such that $P(b) \neq \emptyset$.

Proof We know A is TU $\iff$ $[A, I]$ unimodular.
$P(b) = \{x \in \mathbb{R}^n \mid Ax \leq b\}$. Extreme points of $P(b)$ are in 1-1-correspondence with extreme points in $\{(x, y) \mid Ax + Iy = b, y \geq 0\}$. From previous proof result, this result follows.

Theorem - combination property $A \in \mathbb{Z}^{m\times n}$ is TU $\leftrightarrow$ for every subset J of $\{1, \ldots, n\}$ we can partition J into J_1 and J_{-1} such that $\mid \sum_{j\in J} A_{ij} - \sum_{j\in J_{-1}} A_{ij} \mid = 0, -1, +1$, i.e., result vector with entries equals to $0, +1 and -1$ $\forall i \in \{1, \ldots, m\}$.

Proof:

- "A is TU $\rightarrow$ ..."

  Let $J \subseteq \{1, \ldots, n\}$. Define $d \in \{0, 1\}^n$, $d_j = 1 \forall j \in J$, $d_i = 0 \forall i \notin J$.
  $f \in \mathbb{R}^n$, $\lfloor f \rfloor = \begin{pmatrix} \lfloor f_1 \rfloor \\ \ldots \\ \lfloor f_n \rfloor \end{pmatrix}$, $\lceil f \rceil = \begin{pmatrix} \lceil f_1 \rceil \\ \ldots \\ \lceil f_n \rceil \end{pmatrix}$.
  $P = \{x \in \mathbb{R}^n \mid 0 \leq x \leq d, \lfloor \frac{1}{2}Ad \rfloor \leq Ax \leq \lceil \frac{1}{2}Ad \rceil\}$, $\frac{d}{2} \in \mathcal{P} \rightarrow P \neq \emptyset$, $\mathcal{P}$ is bounded $\rightarrow \mathcal{P}$ has integral extreme point $x \in \{0, 1\}^n$ $(0 \leq x \leq d)$.
  Consider $y = d - 2x$, $y_i = 0 \forall i \notin J \rightarrow d = 0$, and x is also 0. $y_j \in \{-1, +1\} \forall j \in J$.
  Show: $Ay \in \{-1, 0, 1\}$.

  - Suppose $(Ad)_i = 2k \rightarrow (Ay)_i = 0$
  - Suppose $(Ad)_i = 2k + 1 \rightarrow$ then in $\mathcal{P}$: $k \leq (Ax)_i \leq k + 1$.

$(Ay)_i = (Ad)_i - 2(Ax)_i = 2k+1-2\begin{cases} k & \text{if } (Ax)_i = k \\ k+1 & \text{if } (Ax)_i = k+1 \end{cases}$ Define $J_1 = \{j \in J : y_j = 1\}$, $J_2 = \{j \in J, y_j = -1\}$, $J_1 \cap J_2 = \emptyset$, $J_1 \cup J_2 = J$ and $|(Ay)_i| = |\sum_{j\in J_1} A_{ij} - \sum_{j \in J_2} A_{ij}| \leq 1$ $\forall i = 1, \dots, m$.

- Conversely:

  Let B be a $k \times k$ nonsingular square submatrix of A, $r = det[B]$. Goal: $|r| = 1$.
  Induction on k:

  - $k = 1$: trivial, "partition property".
  - $k > 1$:

    $A_{ij} \in \{-1, 0, +1\}$, $B_{ij} \in \{-1, 0, 1\}$. $(B^{-1})_{ij} = \frac{det[B^{ij}]}{r}$. $B^{ij}_{\cdot l} = B_{\cdot l} \forall l \in \{1, \dots, m\} \setminus \{i\}$. $B^{ij}_{\cdot i} = e_j$.
    From our hypothesis of induction using the fact $B^{ij}_{\cdot i} = e_j$, $det[B^{ij}] \in \{0, \pm 1\}$. Let $\bar{A}$ be the matrix with entries $det[B^{ij}]$.
    $B^{-1} = \frac{\bar{A}}{r}$, $\bar{A}_{ij} \in \{-1, 0, 1\}$.
    $B \times \bar{A}_{\cdot 1} = r \times e_1$ (*)
    Let $J = \{i : \bar{A}_{i1} \neq 0\}$, $J_1' = \{i \in J, \bar{A}_{i1} = 1\}$. $B \times \bar{A}_{\cdot 1} = \sum_{j \in J'} b_{ij} - \sum_{j \in J \setminus \{J_1'\}} b_{ij} = 0$ $\forall i = 2, \dots, k\} \rightarrow |\{j \in J \mid b_{ij} \neq 0\}|$ is even. $\rightarrow \forall (J_1, J_2)$ a partition of J such that $\underbrace{|\sum_{j \in J_1} b_{ij} - \sum_{j \in J_2} b_{ij}|}_{=0 \forall i \in \{2, \dots, k\}} \leq 1 \ \forall i \in \{2, \dots, k\}$.
    From assumptions, there exists such a partition of J into J_1 and J_2 such that: $|\sum_{j \in J_1} b_{ij} - \sum_{j \in J_2} b_{ij}| = 0, \forall i = 2, \dots, k$, $|\sum_{j \in J_1} b_{ij} - \sum_{j \in J_2} b_{ij}| \leq 1$
    $\rightarrow \exists z \in \{0, \pm 1\}^k$ such that $B_z = \lambda e_1$, where $\lambda \in \{+1, -1\}$.
    From (*): $B\bar{A}_{\cdot 1} = re_1 \rightarrow \underbrace{\bar{A}_{\cdot 1}}_{\in \{-1, 0, +1\}^k} = \frac{r}{\lambda} \underbrace{z}_{\in \{-1, 0, 1\}^k} \rightarrow |r| = 1$.

supplementary: sections 19.1 and 19.2 from [sch86]

Applications of Total Unimodularity

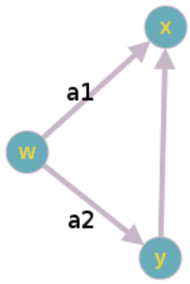
A important reminder: Theorem: $A \in \{0, \pm 1\}^{m \times n}$ is TU iff $\forall J \subseteq [n]$ (or $J \subseteq [m]$ because A^T is also TU) $\exists$ partition $J = J_1 \cup J_2$ such that $\sum_{j \in J_1} A_{ij} - \sum_{j \in J_2} A_{ij} \in \{0, \pm 1\}$ iff $\max\{c^T x \mid Ax \leq b\}$ has an **integral** optimal solution for every $b \in \mathbb{Z}^n$, wherever an optimal solution exists, iff A^T is TU.

Definition - Digraph Let V be a finite set and $A \subseteq V \times V$. Then $D = (V, A)$ is a digraph (directed graph).

Definition - Node-arc incidence matrix The node-arc incidence matrix of $D = (V, A)$ is the matrix $M \in \{0, \pm 1\}^{|V| \times |A|}$, where $M_{ia} = \begin{cases} 1 & \text{if } (i, j) = a \in A \\ -1 & \text{if } (j, i) = a \in A \\ 0 & \text{else} \end{cases}$

Definition - Node-edge incidence matrix The node-edge incidence matrix of an undirected graph $G = (V, E)$ is $M \in \{0, 1\}^{|V|\times|E|}$, where $M_{ve} = \begin{cases} 1 & \text{if } e = (v, w) \in E \\ 0 & \text{else} \end{cases}$

Definition - Edge sets δ^+, δ^- and δ Let $D = (V, A)$ and Let $W \subseteq V$.
$\delta^+(W) = \{(i, j) \in A, i \in W, j \notin W\}$, i.e., all the arcs leaving W.
$\delta^-(W) = \{(i, j) \in A, i \notin W, j \in W\}$, i.e, all the arcs entering W.
Let $G = (V, E)$ and $i \in V$. Then, $\delta(i) = \{(i, j) \in E\}$.



Example: $\delta^+(w) = \{a1, a2\}$.
$\delta^-(w) = \emptyset$.

Theorem: The following matrices are TU:

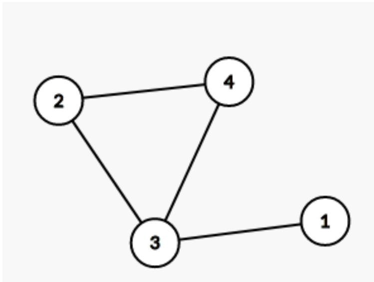
1. The node-arc incidence matrix of a digraph
2. The node-edge incidence matrix of a bipartite undirected graph
3. An interval matrix, i.e, a $\{0, 1\}$ matrix where in each row, the 1's are consecutives.

Proof:

1. Let M be a node-arc incidence matrix of a digraph $D = (V, A)$. Every column of M has one 1's and one -1's. Let $J \subseteq [|V|]$ be a subset of the rows. Let $J_1 = J$ and $J_2 = \emptyset$. Thus, $\sum_{i\in J_1} M_{ij} - \underbrace{\sum_{i\in J_2} M_{ij}}_{=0} = \sum_{i\in J_1} M_{ij} \in \{0, \pm1\}^{|A|}$.
2. Let M be a node-edge incidence matrix of a bipartite graph $G = (V, E)$. Every column of M has two 1's. Let $J \subseteq [|V|]$ be a subset of the rows. Since G is bipartite, $V = V_1 \cup V_2$. Let $J_1 = J \cap V_1$ and $J_2 = J \cap V_2$. Thus, $\sum_{i\in J_1} M_{ij} - \sum_{i\in J_2} M_{ij} \in \{0, \pm1\}^{|A|}$.
3. exercise

Maximum stable set problem

Given $G = (V, E)$, find largest $W \subseteq V$ st no two vertices in W are adjacent (no edges between them).



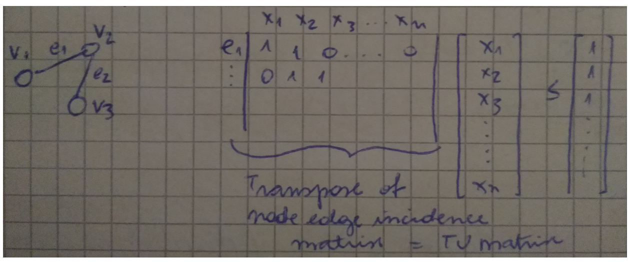
Example: Some stable sets are: $\{v_1\}$, $\{v_1, v_2\}$, $\{v_1, v_4\}$, and the maximum for this graph contains two vertices.

We can model this problem as an ILP If we have a cost function $c \in \mathbb{R}^V$, the objective function is written as: $\sum_{v \in V} x_v c_v$, the constraints are kept the same as in the ILP:

$$\begin{aligned} \max \quad & \sum_{v \in V} x_v \\ \text{st} \quad & x_v + x_w \leq 1, \ \forall \{v, w\} \in E \\ & x_v \in \{0, 1\} \ \forall v \in V \end{aligned}$$

Proposition: If G is bipartite, then the optimal solution to $\max\{\sum_{v \in V} x_v \mid x_v + x_w \leq 1, x_v \in [0, 1]\}$ is integral.

Proof This follows since the constraint matrix is the node-edge incidence matrix of an undirected bipartite graph.
Let A be the node-edge incidence matrix of a bipartite graph G. There is a $1-1$ correspondece between the integral points in $P = \{x \in \mathbb{R}^n_+ \mid A^T x \leq e\}$ where $e = 1^{|E|}$ and the stable sets of G. Therefore, optimizing over P, which is defined by a TU matrix, yields an integral solution.



Definition - (s-t)-flow Let $D = (V, A)$ be a digraph. For each $a \in A$, $c_a \in \mathbb{R}_+$ is the *arc capacity*. Let $s, t \in V$ with $s \neq t$. A function $x : A \mapsto \mathbb{R}$ is an *(s-t)-flow* if:

1. flow conservation constraint: $\sum_{a\in\delta^+(v)} x_a = \sum_{a\in\delta^-(v)} x_a$, $\forall v \in V \setminus \{s, t\}$, i.e., the amount of flow arriving in an intermediate node must be the same amount that comes out.
2. $0 \leq x_a \leq c_a \ \forall a \in A$

The value of x is $val(x) = \sum_{a\in\delta^+(s)} x_a - \sum_{a\in\delta^-(s)} x_a$.

check figure on notes

Theorem: If $c_a \in \mathbb{Z}_+$ for all $a \in A$, then the max value of an (s-t)-flow can be chosen to be integral.

Proof: Let M be the node-arc incidence matrix of D. Define $b \in \mathbb{R}^{|V|}$ as $b_i = \begin{cases} 0 \text{ if } i \in V \setminus \{s,t\} \\ ||c||_1 \text{ if } i \in \{s,t\} \end{cases}$

A max (s-t)-flow is the solution to $\max\{\sum_{e\in\delta^+(s)} x_e - \sum_{e\in\delta^-(s)} x_e$ st $-b \leq Mx \leq b$ and $0 \leq x \leq c\}$. As M is TU, the optimal solution to the problem is integral.

Definition - cut Let $D = (V, A)$ and $W \subseteq V$. The set $\delta^+(W)$ is a cut (induced by W). An (s-t)-cut satisfies $s \in W$ and $t \notin W$. Given $c \in \mathbb{R}_+^{|A|}$, the capacity of $\delta^+(W)$ is $c(\delta^+(W)) = \sum_{a\in\delta^+(W)} c_a$.

example on notes

Lemma: Given $D = (V, A)$ and $s \neq t \in V$, let x be an (s-t)-flow. For every $W \subseteq V$ st $s \in W$, $t \notin W$, $val(x) = \sum_{a\in\delta^+(W)} x_a - \sum_{a\in\delta^-(W)} x_a$. Intuitively, $val(x)$ is (flow out of W) - (flow into W).

Proof proof requires some manipulation of definition

prove!

Lemma: Let x be an (s-t)-flow and $\delta^+(W)$ an (s-t)-cut. Then $val(x) \leq c(\delta^+(W))$, i.e., the amount of the flow is at most the amount of the cut.

Proof: Let x be an (s-t)-flow. Let $W \subseteq V$ be such that $s \in W$ and $t \notin W$. The capacity of the (s-t)-cut induced by W is $c(\delta^+(W))$. $val(x) = \sum_{a\in\delta^+(W)} x_a - \sum_{a\in\delta^-(W)} x_a$, each x is bounded by 0 and c_a, then $val(x) \leq \sum_{a\in\delta^+(W)} c_a - \sum_{a\in\delta^-(W)} 0 = c(\delta^+(W))$.

Theorem: Let $D = (V, A)$, $s \neq t \in V$, $c \in \mathbb{R}_+^{|A|}$. The max value of an (s-t)-flow is equal to the min capacity of an (s-t)-cut.

Proof: The max flow can be solved using:

$$\begin{aligned} \max\ \ & z \\ \text{st}\ \ & x(\delta^-(v)) - x(\delta^+(v)) = 0,\ \forall v \in V \setminus \{s,t\} \\ & x(\delta^-(s)) - x(\delta^+(s)) + z = 0 \\ & x(\delta^-(t)) - x(\delta^+(t)) - \bar{z} = 0 \\ & x_a \le c_a,\ \forall a \in A \\ & x_a \ge 0,\ \forall a \in A \end{aligned}$$

The dual LP is

$$\begin{aligned} \min\ \ & \sum_{a\in A} c_a y_a \\ \text{st}\ \ & y_a + z_v - z_u \ge 0,\ \forall a = (u,v) \in A \\ & z_s = 0 \\ & z_t = 1 \\ & y_a \ge 0,\ \forall a \in A \end{aligned}$$

let x^* be an optimal solution to LP (max). From previous lemma, it is enough to find an (s-t)-cut satisfying $val(x^*) = c(\delta^+(W))$. Let (y^*, z^*) be an optimal dual solution and set $w = \{u \in V \mid z_u^* > 0\}$. If $a = (u,v) \in \delta^+(W)$ then $z_u^* > 0$, $z_v^* \le 0$.

$y_a^* > 0$. By complementary slackness $x_a^* = c_a$ if $a = (u,v) \in \delta^-(W)$, then $y_a^* + z_v - z_u > 0$, so, $x_a^* = 0$. Thus, $val(x) = \sum_{a\in\delta^+(W)} x_a - \sum_{a\in\delta^-(W)} x_a = \sum_{a\in\delta^+(W)} c_a - \sum_{a\in\delta^-(W)} 0 = c(\delta^+(W))$.

Algorithms for hard problems

knapsack problem

Definition - knapsack problem Given a set $C = \{c_1, \ldots, c_n\} \subseteq \mathbb{Z}_+ \setminus \{0\}$ (values) and $A = \{a_1, \ldots, a_n\} \subseteq \mathbb{Z}_+ \setminus \{0\}$ (weights), $a_0 \in \mathbb{Z}_+ \setminus \{0\}, a_0 \ge a_i \forall i$.

$\gamma(k, a_0) = \max\{\sum_{i=1}^k c_i x_i \mid \sum_{i=1}^k a_i x_i \le a_0, x \in \{0,1\}^k\}$ is a (binary) knapsack problem. (If $x \in \mathbb{Z}_+^k$, then it is a integer knapsack problem).

Dual: $\alpha(k, c_0) = \min\{\sum_{i=1}^k a_i x_i \mid \sum_{i=1}^k c_i x_i \ge c_0, x \in \{0,1\}^k\}$

A k-cursive algorithm For all $a \in \{1, \ldots, \sum_{i=1}^n a_i\}$, initialize $\gamma(1, a_0)$:

$$\gamma(1, a_0) = \begin{cases} 0, a_1 > a_0 \\ c_1, a_1 \le a_0 \end{cases}$$

For $k = 2, \ldots, n$ and for $a = 1, \ldots, \sum_{i=1}^n a_i$: $\gamma(k, a) = \max\{\gamma(k-1, a), \gamma(k-1, a - a_k) + c_k\}$

The number of iterations for a given a_0 is $\mathcal{O}(a_{max} n^2)$. Since $a_0 \le n a_{max}$, $\mathcal{O}(n \times a_0)$.

For the dual knapsack problem: for all $c \in \{1, \ldots, \sum_{i=1}^{n} c_i\}$, initialize $\alpha(1, c_0)$:

$$\alpha(1, c_0) = \begin{cases} +\infty, c_1 < c_0 \\ a_1, \text{ if } 1 \leq c_0 \leq c_1 \\ 0, \text{ if } c_0 = 0 \end{cases}$$

For $k = 2, \ldots, n$ and for $c = 1, \ldots, \sum_{i=1}^{n} c_i$: $\alpha(k, c) = \min\{\alpha(k-1, c), \alpha(k-1, c-c_k) + a_k\}$

Corollary The knapsack problem can be solved in time $\mathcal{O}(n \times a_0)$ and the dual version in time $\mathcal{O}(n \times c_0) = \mathcal{O}(n^2 \times max\{c_i : i = \{1, \ldots, n\}\})$

Notation For numbers $d_1, \ldots, d_k$, $d_{max} = max\{d_i : i = \{1, \ldots, k\}\}$

Connections primal/dual knapsack

Lemma $\alpha(k, \gamma(k, a_0) \leq a_0$ and $\gamma(k, \alpha(k, c_0) \geq c_0$

Proof: Let $x^* \in \{0,1\}^k$ attain optimal value for $\gamma(k, a_0)$. Then:

- (i) $\sum a_i x_i^* \leq a_0$
- (ii) $\sum c_i x_i^* = \gamma(k, a_0)$

Consider the dual: $\min\{\sum_{i=1}^{k} a_i x_i \mid \sum_{i=1}^{k} c_i x_i \geq \gamma(k, a_0)\}$, x^* is feasible for the dual, $\leq \sum a_i x_i^* \leq a_0$.

Lemma Let $\mathbb{A}$ be an algorithm for computing $\alpha(k, c_0) \forall k$ and c_0. With "polynomial" many calls of $\mathbb{A}$ we can compute $\gamma(k, a_0)$.

Proof: $\gamma(k, a_0) = \max c^T x, a^T x \leq a_0, x \in \{0,1\}^k$. Assume $a_0 \geq a_{max}$:

$$\underbrace{0}_{\underline{\mu}} \leq \gamma(k, a_0) \leq \underbrace{\sum_{i=1}^{k} c_i}_{\bar{\mu}}.$$

Consider $\mu = \frac{\bar{\mu} + \underline{\mu}}{2}$. Compute $\alpha(k, \mu)$, then two things ca happen:

- (i) the result is greather than $a_0 \rightarrow \bar{\mu} = \mu$
- (ii) the result is less or equal to $a_0 \rightarrow \underline{\mu} = \mu$

  The number of steps that takes until $\underline{\mu} = \bar{\mu}$ is $log(k \times c_{max})$.

Remark $\alpha(k, \mu)$ can be computed by finding a shortest path in a digraph $D = (V, A)$, $V = \{0, \ldots, n\} \times \{0, \ldots, \sum_1^k c_i\}$. The arc A is of the type $((i, \sigma'), (i+1, \sigma)) \iff \sigma - \sigma' \in \{0, c_i + 1\}$ of weight 0 (if $\sigma - \sigma' = 0$) or $a_i + 1$ (if $\sigma - \sigma' = c_i + 1$). Find a shortest path from $(0,0)$ to a node (k, σ), where $\sigma \geq \mu$. This is a "flow problem" in an "exponentially" large graph.

Two approximation results

Theorem: Let $\frac{c_1}{a_1} \geq \frac{c_2}{a_2} \geq \cdots \geq \frac{c_n}{a_n}$. With $\mathcal{O}(n)$ comparisons we can determine $\bar{x} \in \{0,1\}^n$, $a^T\bar{x} \leq a_0$ (feasible) and $c^T\bar{x} \geq \frac{1}{2}\gamma(n, a_0)$.

Proof: Let s be such that $\sum_1^s a_i \leq a_0$ but $\sum_1^s a_i + a_s + 1 > a_0$ wlog, $a_i \leq a_0$, $\forall is \geq 1$.
$\gamma(n, a_O) \leq \sum_1^s c_i + c_{s+1}\frac{a_0 - \sum_1^s a_i}{a_{s+1}} \leq \sum_1^s c_i + c_{s+1} = \underbrace{c^T x_1}_{\text{all elements until } s} + \underbrace{c^T x_2}_{\text{just } s+1 \text{ element, since } a_0 \geq a_i}$.
Where x_1 and x_2 are feasible $0/1$-solutions.
$c^T x_1 + c^T x_2 \leq 2\max\{c^T x_1, c^T x_2\}$.

FPTAS: Fully polynomial time approximation scheme A family of algorithms $\mathcal{A}_\varepsilon$ such that $\mathcal{A}_\varepsilon$ finds a feasible $0/1$-solution for knapsack of value $c(\bar{x}) \geq (1-\varepsilon) \times OPT$ (max problem) or $c(\bar{x}) \leq (1+\varepsilon) \times OPT$ (min problem), where OPT is the optimal objective value, in running time polynomial in $\frac{1}{\varepsilon}$ and in the binary encoding of the instance.

Theorem - FPTAS Let $\varepsilon > 0$. In time $\mathcal{O}(n^3\frac{1}{\varepsilon}log(\frac{n^2}{\varepsilon}))$ we can compute a feasible solution $\bar{x} \in \{0,1\}^n$, $a^T\bar{x} \leq a_0$, such that $c^T\bar{x} \geq (1-\varepsilon)\gamma(n, a_0)$ ($\bar{x}$ is not far away from the optimal solution).

Proof: Let $\Delta = \frac{\varepsilon c_{max}}{n} \to \varepsilon = \frac{\Delta n}{c_{max}} \to c_{max} = \frac{\Delta n}{\varepsilon}$.
$\frac{c^T}{\Delta} \geq c' = (\lfloor\frac{c_1}{\Delta}\rfloor, \ldots, \lfloor\frac{c_n}{\Delta}\rfloor) \geq (\frac{c_1}{\Delta} - 1, \ldots, \frac{c_n}{\Delta} - 1)$ (*)
$c'_{max} \leq \frac{c_{max}}{\Delta} = \frac{n}{\varepsilon}$
Compute an optimal solution for $\max\{c'^T x \mid a^T x \leq a_0, x \in \{0,1\}^n\}$. This can be done with the lemma before (α, γ- relation) in time $\mathcal{O}(n^3\frac{1}{\varepsilon}log(\frac{n^2}{\varepsilon}))$. Let $\bar{x}$ be this solution. Let x^* be an optimal solution for $\max\{c^T x \mid a^T x \leq a_0, x \in \{0,1\}^n\} \to c^T x^* = \gamma(n, a_0)$.
$(1-\varepsilon)\gamma(n, a_0) - \varepsilon\gamma(n, a_0) = \gamma(n, a_0) - \frac{\Delta n}{c_{max}}\underbrace{\gamma(n, a_0)}_{\geq c_{max}} \leq \gamma(n, a_0) - \Delta n = c^T x^* - \Delta n \leq c^T x^* - \Delta\mathbb{1}^T x^* \to$
$\Delta(\frac{c_1}{\Delta} - 1, \ldots, \frac{c_n}{\Delta} - 1)x^* \underbrace{\leq}_{(*)} \Delta c'^T x^* \leq \Delta c^T\bar{x} \leq c^T x$.

Modeling with Discrete Variables

Definition: Assume that you are given $F \subseteq \mathbb{Z}^n \times \mathbb{R}^d$ such that $P = conv(F)$.

1. A model for F is a system of linear inequalities $Q = \{x \in \mathbb{R}^s \times \mathbb{R}^t : Ax \leq b\}$ with rational data A and b together with a rational linear map $\phi : \mathbb{R}^{s+t} \mapsto \mathbb{R}^{n+d}$ such that $\phi(Q) \cap (\mathbb{Z}^n \times \mathbb{R}^d) = F$.

2. For polyhedra A and B in $\mathbb{R}^n$, let $d(A, B)$ be a distance function which fulfills $d(A, A) = 0$ and for polyhedra $A \subseteq B \subseteq C$, we postulate that $d(A, B) \leq d(A, C)$. Then, we define the quality of a model Q for F as $d(P, Q)$.

3. A model is an ideal formulation for F if $d(P, Q) = 0$.

Set covering, packing and partitioning

Definition: Let M be a finite ground set. $M = \{1, \ldots, m\}$. Given a collection $M_1, \ldots, M_n \subseteq M$, $c_j > 0$ is the weight of subsets M_j, $M_j = \{1, \ldots, n\}$.

Cover: $J \subseteq \{1, \ldots, n\}$ is a cover of M if $\bigcup_{j \in J} M_j = M$.

Packing: $J \subseteq \{1, \ldots, n\}$ is a packing of M if $M_i \cap M_j = \emptyset$, $\forall i \in J, i \neq j$.

Partition: A partition is a subset $J \subseteq \{1, \ldots, n\}$ such that J is a cover and a packing.

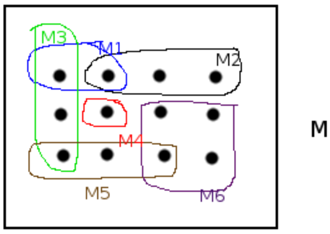


Figure 8: $J = \{2, 5\}$ is a packing. There is no partition on this set.

The optimization problems:

1. Max-weight packing: we want to find a packing of maximum weight.
2. Min-weight cover: we want to find a cover of minimum weight.
3. Min/Max-weight partition: partition with min/max weight.

Goal: "translate" these optimization problems into "polyhedral language". Assume that we are given an objective function $c \in \mathbb{Z}^{|N|}$ with $c_i \geq 0, \forall i \in \{1, \ldots, n\}$.It requires to find a matrix $A \in \mathbb{Z}^{m \times n}$, where m is our ground set and n is the subsets of M. $A_{ij} = \begin{cases} 0, \text{ if } i \notin M_j \\ 1, \text{ if } i \in M_j \end{cases}$.

Introduce variables $x_j \forall j \in N$. $x_j \in \{0, 1\}$. Each $x_j = 1$ means to select the set.

1. Max-weight packing: $\max \sum_{j=1}^{n} c_j x_j$, $Ax \leq \mathbb{1} \in \mathbb{R}^m$
2. Min-weight cover: $\min \sum_{j=1}^{n} c_j x_j$, $Ax \geq \mathbb{1} \in \mathbb{R}^m$
3. Min/Max-weight partition: $\min / \max \sum_{j=1}^{n} c_j x_j$, $Ax = \mathbb{1} \in \mathbb{R}^m$, $0 \leq x \leq \mathbb{1}$, $x \in \mathbb{Z}^n$

For a more general problem, we can consider the right hand side $b \in \mathbb{Z}_+^m$ instead of $\mathbb{1}$.

By defining $y_i = 1 - x_i$, $i \in \{1, \ldots, n\}$ we can transform a covering to a packing problem and vice-versa. if x satisfies $Ax \leq b$ then y satisfies $Ay \leq A - b$, $y \in \{0,1\}^n$.

Formulations

Definition - Mixed Integer Optimization Problem (MIP) Mixed = integer and continuous variables: $\max\{\sum c_i x_i + \sum d_j y_j \mid Ax + Dy \leq b, x \in \mathbb{R}^n, y \in \mathbb{Z}^d\}$. Its **linear relaxation** is defined by replacing $y \in \mathbb{Z}^d$ by $y \in \mathbb{R}^d$.

Example (facility location problem): Customers $i \in \{1, \ldots, n\}$. Potential facilities $j \in \{1, \ldots, m\}$. We have a fixed cost c_j for opening facility j. Variable cost d_{ij} for serving client i from facility j (of course, if it is open).
$x_{ij} \in [0,1]$: percentage of serving client i from j. For all $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m\}$.
$y_j \in \{0,1\}$ $\forall j \in \{1, \ldots, m\}$ indicates if we open j-th facility.
$P_{FL} = \{\sum_{j=1}^n x_{ij} = 1, \forall i \in \{1, \ldots, n\}, x_{ij} \leq y_j \ \forall i, j\}$
Objective: $Z_{FL} = \min \sum_{j=1}^n c_j y_j + \sum_{i,j} d_{ij} x_{ij}$
We could also look at a second version of the problem: $P_{AFL} = \{(x,y) \mid \sum_{j=1}^n x_{ij} = 1, \forall i, \sum_{i=1}^m x_{ij} \leq my_j \forall j, 0 \leq x_{ij} \leq 1, 0 \leq y_j \leq 1\}$
Objective: $Z_{AFL} = \min \sum_{j=1}^n c_j y_j + \sum_{i,j} d_{ij} x_{ij}$, $(x,y) \in P_{AFL}$.
Observe: $P_{AFL} \supseteq P_{FL} \supseteq P_{FL} \cap (\underbrace{\mathbb{R}^{n \times m}}_{x} \times \underbrace{\mathbb{Z}^m}_{y}) = P_{AFL} \cap (\mathbb{R}^{n \times m} \times \mathbb{Z}^m)$
$(x,y) \in P_{FL} \rightarrow x_{ij} \leq y_i \forall i \rightarrow \sum_{i=1}^m x_i j \leq my_j$

Let A and B be two formulations of the same integer optimization problem and P_A and P_B be the feasible sets of the corresponding linear relaxations. We say, that formulation A is at least as strong as formulation B if $P_A \subseteq P_B$

Example (minimum spanning tree): Given $G = (V, E)$, weights $w_e \forall e \in E$, G connected, a forest $F \subseteq E$, (V, F) contains no cycle.

Definition - spanning tree: $T \subseteq E$ is a spanning tree if (V, T) has no cycles and $|T| = n - 1$, $n = |V|$.

Lemma - graph theory: $T \subseteq E$ is a spanning tree $\iff$ $|T| = n - 1$ and (V, T) is connected.

The cost of a spanning tree is the sum of the costs of the edges of the tree.

Two formulations for min-weight spanning trees

Consider new variables $x_{ij} \in \{0,1\} \forall (i,j) \in E$.

Notation: For $S \subseteq V$, $E(S) = \{(i,j) \in E \mid i,j \in S\}$. $\delta(S) = \{(i,j) \in E \mid i \in S, j \notin S\}$, *i.e., a cut comming from* S.

Subtour-elimination formulation (relaxation) $P_{SUB} = \{x \in [0,1]^{|E|} \mid \sum_{e \in E(S)} x_e = n-1, \sum_{e \in E(S)} x_e \leq |S| - 1 \forall \emptyset \neq S \neq V,\ S \subseteq V\}$

Optimization problem 1: $Z_{SUB} = \min\{x \in P_{SUB} \cap \mathbb{Z}^n \sum_{e \in E} w_e x_e\}$

Cut set formulation: $P_{cut} = \{x \in [0,1]^{|E|} \mid \sum_{e \in E(S)} x_e = n-1, \sum_{e \in \delta(S)} x_e \geq 1,\ \forall \emptyset \neq S \neq V,\ S \subseteq V\}$.

Optimization problem 2: $Z_{cut} = \min\{x \in P_{cut} \cap \mathbb{Z}^n \sum_{e \in E} w_e x_e\}$

Theorem:

1. $P_{SUB} \subseteq P_{cut}$
2. In general, $P_{SUB} \neq P_{cut}$ and P_{cut} may have non integral extreme points.

Proof:

1. x in P_{SUB}. Show $S \subseteq V$, $S \neq 0$, $S \neq V$, $\sum_{e \in \delta(S)} x_e \geq 1$.
   $E = E(S) \cup \delta(S) \cup E(V\S)$. $x \in P_{SUB} \rightarrow \sum_{e \in E(S)} x_e \leq |S| - 1$. $\sum_{e \in E(V\S)} x_e \leq |V| - |S| - 1 = n - |S| - 1 \rightarrow n - 1 = \sum_{e \in E(S)} x_e + \sum_{e \in \delta(S)} x_e + \sum_{e \in E(V\S)} x_e \rightarrow \sum_{e \in \delta(S)} x_e = n - 1 - \underbrace{(\sum_{e \in E(S)} x_e)}_{\leq |S|-1} - \underbrace{(\sum_{e \in E(V\S)} x_e)}_{\leq n-|S|-1} \geq n - 1 - (|S| - 1) - (n - |S| - 1) = -|S| + |S| + 1 = 1$
2.

From Linear to Integer Optimization

The integer convex hull of a polyhedron

Theorem for integer points in polyhedra Note: The convex hull of the integer points of a polyhedron is also a polyhedron.

Let $P = \{x \in \mathbb{R}^n \mid Ax \leq b\}$ and $\mathcal{F} = P \cap \mathbb{Z}^n$.

There exists a finite set $T \subseteq \mathcal{F}$ such that
$\mathcal{F} = \{x \in \mathbb{R}^n : x = \sum_{t \in T} \lambda_t t + \sum_{e \in E} \mu_e e,\ \lambda_t \in \mathbb{Z}_+ \forall t \in T, \mu_e \in \mathbb{Z}_+ \forall e \in E, \quad \underbrace{\sum_{t \in T} \lambda_t = 1}_{\text{exists only one } \lambda \neq 0 \text{ and } \lambda = 1} \quad \}$

where $E \subseteq \mathbb{Z}^n$ and $\underbrace{\mathcal{C}}_{\text{recession cone}} = \{x \in \mathbb{R}^n \mid Ax \leq 0\} = cone(E) = rec(P) = conv(P \cap \mathbb{Z}^n)$

Proof: $P = conv(V) + cone(E) \to$ the polyhedron can be written as a convex hull of its vertices plus the cone of edges.

$T = \{x \in \mathbb{R}^n \mid x = \sum_{v \in V} \lambda_v v + \sum_{e \in E} \lambda_e e, \sum_{v \in V} \lambda_v = 1, 0 \leq \lambda_v \leq 1 \forall v \in V, 0 \leq \lambda_e \leq 1 \forall e \in E\} \cap \mathbb{Z}^n$

$|T|$ is finite $\to P$ is bounded.

Let $x \in \mathcal{F}$. $\exists \lambda_v \geq 0, \forall v \in V, \lambda_e \geq 0 \forall e \in E, \sum_{v \in V} \lambda_v = 1$ such that $x = \sum_{v \in V} \lambda_v v + \sum_{e \in E} \lambda_e e =$
$\underbrace{\sum_{v \in V} \lambda_v v + \sum_{e \in E} (\lambda_e - \lfloor \lambda_e \rfloor) e}_{(**)} + \underbrace{\sum_{e \in E} \lfloor \lambda_e \rfloor e}_{(*)}$

x is an integer. $x - (*)$ is an integer, so $(**)$ is also integer ($\in \mathbb{Z}^n \cap T$).

Theorem: $P = \{x \in \mathbb{R}^n \mid Ax \leq b\} \neq \emptyset$. $\mathcal{F} = P \cap \mathbb{Z}^n$. $conv(\mathcal{F})$ **is a polyhedron** From previous theorem, $\exists T \subseteq \mathcal{F}, |T|$ is finite such that $\forall x \in \mathcal{F}, \exists t_x \in T$ and multipliers $\mu_e \in \mathbb{Z}_+$ such that $x = t_x + \sum_{e \in E} \mu_e e$.

We show that: $conv(\mathcal{F}) = conv(T) + cone(E)$ **The recession cone is the same for the IP and LP.**

Proof: direction "$\subseteq$" Let $x^1, \dots, x^k \in \mathcal{F}$ and $\lambda_1, \dots, \lambda_k \geq 0, \sum \lambda_i = 1$.
$\forall i$, let $t_{x^i} \in T$ and $\mu_e^i \in \mathbb{Z}_+$, $x^i = t_{x^i} + \sum_{e \in E} \mu_e^i e$
$\sum_{i=1}^k \lambda_i x^i = \sum_{i=1}^k \lambda_i (t_{x^i} + \sum_{e \in E} \mu_e^i e) = \sum_{i=1}^k \lambda_i t_{x^i} + \sum_{i=1}^k \lambda_i \sum_{e \in E} \mu_e^i e$
$= \underbrace{\sum_{i=1}^k \lambda_i t_{x^i}}_{\in conv(T)} + \underbrace{\sum_{e \in E} (\underbrace{\sum_{i=1}^k \lambda_i \mu_e^i}_{\geq 0}) e}_{\in cone(E)}$

Proof: direction "$\supseteq$" Let $x \in conv(T) + cone(E)$, then x is a convex combination of integer points in P.

$x = \sum_{t \in T} \lambda_t t + \sum_{e \in E} \mu_e e$, $\mu_e \geq 0\ \forall e$, $\lambda_t \geq 0\ \forall t$, $\sum_t \lambda_t = 1$
Let $\{t_1, \dots, t_k\} \subseteq T$ such that $\lambda_{t_i} \geq 0$
$x = \sum_{i=1}^k \lambda_i t_i + \sum_{e \in E} \mu_e e = \sum_{i=2}^k \lambda_i t_i + \lambda_1 (t_1 + \sum_{e \in E} \frac{\mu_e}{\lambda_1} e)$.

Claim: Suppose $s \in conv(P \cap \mathbb{Z}^n), e \in E$ **and let** $\mu_e > 0 \to s + u_e e \in conv(P \cap \mathbb{Z}^n)$

With the claim, $t_1 + \sum_{e \in E} \frac{\mu_e}{\lambda_1} e \in conv(P \cap \mathbb{Z}^n)$ and $t_2, \dots, t_k \in conv(P \cap \mathbb{Z}^n) \to \lambda_1 (t_1 + sum_{e \in E} \frac{\mu_e}{\lambda_1} e + \sum_{i=2}^k \lambda_i t_i \in conv(P \cap \mathbb{Z}^n)$.

Proof of the claim: $s \in conv(P \cap \mathbb{Z}^n), e \in E, \mu > 0$. $s + \mu e \in conv(P \cap \mathbb{Z}^n)$ wlog $\mu \notin \mathbb{Z}$
Suppose $s = \underbrace{\frac{1}{2}s_1}_{\in P\cap\mathbb{Z}^n} + \underbrace{\frac{1}{2}s_2}_{\in P\cap\mathbb{Z}^n}$. $s_i + e \in P \cap \mathbb{Z}^n \forall i$, $s = \frac{1}{2}(s_1 + e) + \frac{1}{2}(s_2 + e)$.
$s + \mu e = (\mu e - \lfloor \mu e \rfloor)(s + \lceil \mu e \rceil e) + (\lceil \mu e \rceil - \mu e)(s + \lfloor \mu e \rfloor e) \to \lceil \mu e \rceil - \lfloor \mu e \rfloor = 1$.

Cutting plane algorithm

The second theorem allow us to design a "cutting plane algorithm".
We started with polyhedron $P^0 = P = \{x \in \mathbb{R}^n, Ax \leq b\}$. Task: $\max c^T x, x \in P \cap \mathbb{Z}^n$.
For $i = 0, \dots, N$:

1. Let x^* attains the $\max\{c^T x \mid x \in P^i\}$
2. if $x^* \in \mathbb{Z}^n$: problem solved.
3. Otherwise, determine hyperplane $a^T x = \alpha, a \in \mathbb{Q}^n$ (*), $\alpha \in \mathbb{Q}$ such that:
   $\forall x \in P^i \cap Z^n, a^T x \leq \alpha$
   $a^T x^* > \alpha$
4. $P^{i+1} = P^i \cap \{x \mid a^T x \leq \alpha\}$

Notation: an inequality in (*) is called a cutting plane. $a^T x \leq \alpha$: find a hyperplane where all the integer points lies on one side and x in the other side.

Chvátal-Gomory cuts

Questions:

Can we find cutting planes "cheaply"? $\min\{c^T x \mid Ax = b, x \in \mathbb{Z}^n, x \geq 0\}$.
Consider LP-optimal solution $x^* = (x_B^*, x_N^* = 0)$ with basis B, $x_B^* = A_B^{-1} b \geq 0$. Suppose $x^* \notin \mathbb{Z}^n \to \exists i \in B$ such that $x_i^* \notin \mathbb{Z}^n$. $\bar{b} = A_B^{-1} b$, $\bar{A_N} = A_B^{-1} A_N$. Notice $x_i^* = \bar{b_i}$.
Consider row i: $x_i + \sum_{j \in N} \bar{a_{ij}} x_j = \bar{b_i} \underbrace{\geq}_{x_j \geq 0 \forall j} x_i + \sum_{j \in N} \lfloor \bar{a_{ij}} \rfloor x_j$.
$\forall x \in \mathbb{Z}^n$ such that $Ax = b, x \geq 0$ we know $x_i + \sum_{j \in N} \lfloor \bar{a_{ij}} \rfloor x_j \in \mathbb{Z} \leq \lfloor \bar{b_i} \rfloor < b_i$. $\to x_i^* + \sum_{j \in N} \lfloor \bar{a_{ij}} \rfloor x_j^* = \bar{b_i} > \lfloor \bar{b_i} \rfloor$. (integer gomory cuts)

Why is it atractive to add a cutting plane to an LP-relaxation? After addition of a cutting plane we stay dual feasible. Consider $\min\{c^T x \mid Ax = b, x \geq 0\}$. Consider LP-optimal solution with basis B: primal feasibility $\to A_B^{-1} b = \bar{b} \geq 0$; dual feasibility $\to c^T - c_B^T A_B^{-1} A \geq 0$.
Consider the cutting plane: $a^T x \leq \alpha$, but $a^T x^* > \alpha$. We lose primal feasibility: basis B cannot be maintened for primal feasibility.

Consider expanded system: $Ax = b, a^T x \leq \alpha, x \geq 0 \to (Ax = b, a^T x + z = \alpha, x \geq 0, z \geq 0) = c[x|z] = d$. Consider now a new basis $B = B \cup \{m+1\}$. wlog, $B = \{1, \dots, m\}$. $\bar{B}$ is not primal feasible for $\min\{c^T x \mid Ax = b, a^T x + z = \alpha, x \geq 0, z \geq 0\}$ (*).

Lemma: $\bar{B}$ is dual feasible for (*) Proof: $c_{\bar{B}} = \begin{bmatrix} A_B & 0 \\ a^T B & 1 \end{bmatrix}$; $A_{\bar{B}}^{-1} = \begin{bmatrix} A_B^{-1} & 0 \\ -a^T B A_B^{-1} & 1 \end{bmatrix}$.

$$[c \quad 0] - [c_B^T \quad 0] \begin{bmatrix} A_B^{-1} & 0 \\ -a^T B A_B^{-1} & 1 \end{bmatrix} \times c = [c^T \quad 0] - [c_B^T A_B^{-1} \quad 0] \begin{bmatrix} A & 0 \\ a^T & 1 \end{bmatrix} = [c^T - c_B^T A_B^{-1} A \quad 0] \geq 0$$

Cutting plane principles

Integer rounding

Solve optimization problem $\max\{c^T x \mid Ax \leq b, x \in \mathbb{Z}^n\}$. Find $a^T x \leq \alpha$ (halfspaces) such that $a^T x \leq \alpha$ is true $\forall x \in \mathbb{Z}^n \cap P$. $P = \{x \in \mathbb{R}^n \mid Ax \leq b\}$. $a^T x > \alpha$ for some points in $P \setminus \mathbb{Z}^n$.
Consider $u \geq 0$ st $a^T = u^T A \in \mathbb{Z}^n$. We know $a^T x = u^T Ax \leq u^T b, \forall x \in P$. $\forall x \in P \cap \mathbb{Z}^n$, then $a^T x \leq \lfloor u^T b \rfloor$.

Starting with a polyhedron P, we can define an algebraic operator:
From P define for $A \in \mathbb{Z}^{m \times n}$, $b \in Z^m$:
$P^{(1)} = \{x \in \mathbb{R}^n \mid u^T Ax \leq \lfloor u^T b \rfloor \forall u \geq 0$, st $u^T A \in Z^n\}$. $P^{(1)} \supseteq P \cap \mathbb{Z}^n$, $P^{(1)} \cap \mathbb{Z}^n = P \cap Z^n$ (closure of P operator).

Consider the case: $P = \{x \in \mathbb{R}^n \mid Ax \leq b, x \geq 0\}$.
Take any $u \in \mathbb{R}^n$ then $\sum_{j=1}^n \lfloor u^T A_{\cdot j} \rfloor x_j \leq \lfloor u^T b \rfloor$ is satisfied by all $x \in P \cap \mathbb{Z}^n$ (we can round down for any multiplier u because $x \geq 0$).

$P^{(1)} = \{x \in \mathbb{R}^n \mid \sum_{j=1}^n \lfloor u^T A_{\cdot j} \rfloor x_j \leq \lfloor u^T b \rfloor \forall u \in \mathbb{R}^n\}$. $P^{(1)}$ is a polyhedron and $P^{(1)} \supseteq P \cap \mathbb{Z}^n$ and $P^{(1)} \cap \mathbb{Z}^n = P \cap \mathbb{Z}^n$.

Remark - $P = \{x \mid Ax \leq b, x \geq 0\}$ The closure $P^{(1)} = \{x \mid \sum_{j=1}^n \lfloor u^T A_{\cdot j} \rfloor x_j \leq \lfloor u^T b \rfloor, \forall u \geq 0\}$.

Example 1: add example - check notes

Example 2 - the matching polyhedron (polytime problem): $G = (V, E)$, $M \subseteq E$ is a matching, i.e., $m_i \cap m_j = \emptyset, \forall m_i, m_j \in M, m_i \neq m_j$, i.e., no two edges overlap.
$P = \{x \in [0,1]^{|E|} \mid \sum_{e \in \delta(i)} x_e \leq 1,\ \forall i \in V\}$. $P \cap \mathbb{Z}^n = \{X^M \mid M$matching in $G\}$. $X^M \in \mathbb{R}^{|E|}$, $X_e^M = 1$ if $e \in M, X_e^M = 0$ if $e \notin M$.
Consider $\max\{\sum_{e \in E} c_e x_e \mid x \in P \cap \mathbb{Z}^n\}$ and $\max\{\sum_{e \in E} c_e x_e \mid x \in P\}$.

Consider the graph:

add graph

$\hat{x}$ optimal matching solution: $c^T \hat{x} = 2M + \varepsilon$. So, the two formulations above cannot be the same.

Take all inequalities for $i \in S \subseteq V$, $|S|$ odd. $\frac{1}{2} \sum_{e \in \delta(S)} x_e \leq \frac{1}{2}$, summing them up, you get $\sum_{e \in E(S)} x_e + \frac{1}{2} \sum_{e \in \delta(S)} x_e \leq \frac{|S|}{2}$.
Now, add $-\frac{1}{2} x_e \leq 0, \forall e \in \delta(S)$. We obtain: $\sum_{e \in E(S)} x_e \leq \frac{|S|-1}{2}$ must hold $\forall x \in P \cap \mathbb{Z}^n$. Adding this constraint to the previous polytope it becomes integral.

Mixed Integer Rounding

$\{(x, y) \mid x \in \mathbb{Z}, y \geq 0, y \in \mathbb{R}, x - y \leq b\} = F$. wlog, $b \notin \mathbb{Z}$ (if $b \in \mathbb{Z}$, then, nothing to do). Then $x - \frac{y}{1-f(b)} \leq \lfloor b \rfloor$.

Proof: $F = F_1 \cup F_2$, $F_1 = \{(x, y) \in F \mid x \leq \lfloor b \rfloor\}$, $F_2 = \{(x, y) \in F \mid x \geq \lceil b \rceil\}$. $x - \frac{y}{1-f(b)} \leq \lfloor b \rfloor \iff (1 - f(b))(x - \lfloor b \rfloor) \leq y$.

Lemma of the proof: $x - \frac{y}{1-f(b)} \leq \lfloor b \rfloor$ is satisfied by all $(x, y) \in F$.

Proof: $\forall (x, y) \in F_1$: take $x - \lfloor b \rfloor \leq 0, 0 \leq y$. Multiply inequality by $(1 - f(b))$: $(1 - f(b))(x - \lfloor b \rfloor) \leq 0 \to (1 - f(b))$: $(1 - f(b))(x - \lfloor b \rfloor) \leq y$.
$\forall (x, y) \in F_2$: $x \geq \lceil b \rceil \iff -(x - \lfloor b \rfloor) \leq -1$. Take $-(x - \lfloor b \rfloor) \leq -1$ times $f(b)$ and add $x - y \leq b$. This leads to $-f(b)(x - \lfloor b \rfloor) + x - y \leq \lfloor b \rfloor = b - f(b) \to -f(b)(x - \lfloor b \rfloor) + x - \lfloor b \rfloor \leq y \iff (1 - f(b))(x - \lfloor b \rfloor) \leq y$.
$\to \forall (x, y) \in F$ we have $x - \frac{y}{1-f(b)} \leq \lfloor b \rfloor$.

Theorem (how to apply the principle): Let $F = \{x \in \mathbb{Z}^n \mid x \geq 0, Ax \leq b\}$. For $u \in \mathbb{R}^m_+$ and $x \in F$ we have that $\sum_{j=1}^n (\lfloor u^T A_{\cdot j} + max\{0, \frac{f(u^T A_{\cdot j} - f(u^T b)}{1-f(u^T b)}\} x_j) \leq \lfloor u^T b \rfloor$.

Proof - sketch (how to match a high dimension in a low dimension) $N_1 = \{j \mid f(u^T A_{\cdot j}) \leq f(u^T b)\}$. $N_2 = N \setminus N_1$. $\underbrace{\sum_{j \in N_1} \lfloor u^T A_{\cdot j} \rfloor x_j}_{(1)} + \underbrace{\sum_{j \in N_2} \lceil u^T A_{\cdot j} \rceil x_j}_{(2)} - \underbrace{\sum_{j \in N_2} (1 - f(u^T A_{\cdot j})) x_j}_{(3)} \leq u^T b$ satisfied by all $x \in F$.
$(1) + (2) = w \in \mathbb{Z}$, $(2) - (3) = \sum_{j \in N_2} u^T A_{\cdot j} x_j$, $(3) = z \geq 0$.
We obtain $\{(w, z) \mid w \in \mathbb{Z}, z \geq 0, w - z \leq \lfloor u^T b \rfloor + f(u^T b)\} \to w - \frac{z}{1-f(u^T b)} \leq \lfloor u^T b \rfloor$ is correct and give us the previous formula.

Method of Lift and Project

$P = \{x \in \mathbb{R}^n \mid Ax \leq b\}$ when $0 \leq x_i \leq 1 \forall i = 1, \ldots, n$ are part of $Ax \leq b \to P \cap \mathbb{Z}^n \subseteq \{0, 1\}^n$.

Target: compute the convex hull of $P \cap \mathbb{Z}^n$. Ingredients:

- projection $Q = \{(x, y) \in \mathbb{R}^{n \times d} \mid \tilde{A}x + \tilde{B}y \leq \tilde{b}\}$. $proj_x(Q) = \{x \in \mathbb{R}^n \mid (x, y) \in Q\}$.
- algebraic quadratic operator: $x^2 = x \forall x \in \{0, 1\}$.

Algorithm - Lift and Project: Input: P, $j \in \{1, \ldots, n\}$. Output: polyhedron P_j.

1. Lift: multiply $Ax \leq b$ by x_j and $(1 - x_j)$:

   Consider $x_j Ax \leq bx_j$, $(1 - x_j)Ax \leq b(1 - x_j)$ (*)

   Define variables: $y_{ij} = x_i x_j, \forall i \neq j$.

   Replace the system (*) by the another system where we put y_{ij} in place of $x_i x_j$ and $x_j^2 = x_j$, so our formulation is purely linear (not quadratic anymore).

   Call the resulting polyhedron $L_j(P) \subseteq \mathbb{R}^{n+(n-1)}$.

2. Project:

   Project $L_j(P)$ to the space of x-variables, $P_j = proj(L_j(P))$.

Remark: for $x \in [0,1]$ it holds that $x^2 = x \iff x \in \{0,1\}$. Furthermore, for $y_{kj} = x_k x_j$, where x_k and $x_j \in [0,1]$, it holds that $y_{kj} = 1 \iff x_k = x_j = 1$.

Theorem: $P_j = conv(P \cap \{x \in \mathbb{R}^n \mid x_j \in \{0,1\}) = conv(P \cap \{x \mid x_j = 0\} \cup (P \cap \{x \mid x_j = 1\})$. **$P \neq \emptyset$.**

Proof:

- "$conv(...) \subseteq P_j$"

  Let $\bar{x} \in conv(P \cap \{x \in \mathbb{R}^n \mid x_j \in \{0,1\})$ st $\bar{x_j} \in \{0,1\}$ (we just need to show for 0 and 1 because P is a polyhedron and thus, convex). Define $\bar{y_{ij}} = \bar{x_i}\bar{x_j}$. $\bar{x}$ satisfies (*), $\bar{x_j}^2 = \bar{x_j}$. So we know $(\bar{x}, \bar{y}) \in L_j(P) \rightarrow \bar{x} \in P_j$

- "$P_j \subseteq conv(...)$"

  We distinguish three cases:

  - $P \cap \{x \in \mathbb{R}^n \mid x_j = 0\} = \emptyset$

    We need to prove that P lies on $x_j = 1$. By Farka's lemma: $\exists u \geq 0$ st $u^T A = -e_j$, $-\varepsilon = u^T b < 0$, where $\varepsilon > 0$. In particular, all x satisfying (*), satisfy $(1 - x_j)u^T Ax \leq u^T b(1 - x_j) \rightarrow -(1 - x_j)x_j \leq -\varepsilon(1 - x_j) \rightarrow -x_j + \underbrace{x_j^2}_{=x_j} \leq -\varepsilon(1 - x_j) \rightarrow \forall (x, y) \in L_j(P)$:

    $0 \leq -\varepsilon(1 - x_j) \rightarrow x_j \geq 1$. $\rightarrow \forall x \in P_j$, then x_j is valid.

  - $P \cap \{x \in \mathbb{R}^n \mid x_j = 1\} = \emptyset$

    similar as 01, exercise!

  - $P \cap \{x \mid x_j = 0\} \neq \emptyset$ and $P \cap \{x \mid x_j = 1\} \neq \emptyset$

    Take any inequality $a^T x \leq \alpha$ valid for $conv(P \cap \{x \in \mathbb{R}^n \mid x_j \in \{0,1\}\})$ and show it is valid for P_j.

    Claim: $\exists \mu < 0, \lambda < 0$ st $a^T x + \mu x_j \leq \alpha$ (1) and $a^T x + \lambda(1 - x_j) \leq \alpha$ are valid for P.

    Proof of the claim: Let V be all extreme points in P, st $0 < v_j < 1$ (points 0 and 1

already proved). We have finite points since P is bounded. Define $\mu = \min\{\frac{-a^T v+\alpha}{v_j} \mid v \in V\}$, $\lambda = \min\{\frac{-a^T v+\alpha}{1-v_j} \mid v \in V\}$. From the claim, every x satisfying (*) satisfies $(1-x_j)(a^T x + \mu x_j) \leq \alpha(1-x_j)$, $x_j(a^T x + \lambda(1-x_j)) \leq \alpha x_j$.
All points x in (*) satisfies also the sum: $a^T x + (\lambda+\mu)\underbrace{x_j(1-x_j)}_{=x_j-x_j^2=0} \leq \alpha \to a^T x \leq \alpha$ is satisfied by all x in P_j.

Definition: For $i_1,\dots,i_t \in N$, all distinct, $P_{i_1},\dots,P_{i_t} = (P_{i_1,\dots,i_{t-1}})_{i_t}$.

Theorem: For any sequence $i_1,\dots,i_t \in N$, all distinct, then $P_{i_1,\dots,i_t} = conv(P \cap \{x \in \mathbb{R}^n \mid x_k \in \{0,1\}$, $\forall k \in \{i_1,\dots,i_t\}\})$

Proof: Induction on t. We know it is true for $t=1$, from theorem before. Let $Q = \{x \in \mathbb{R}^n \mid x_k \in \{0,1\} \forall k \in \{i_1,\dots,i_{t-1}\}\}$. Assume theorem is true for sequence $i_1,\dots,i_{t-1}$.
$P_{i_1},\dots,P_{i_t} = (P_{i_1,\dots,i_{t-1}})_{i_t} \underbrace{=}_{\text{from hypothesis}} conv(P \cap Q)_{i_t} \underbrace{=}_{\text{from previous theorem}} conv(conv(P \cap Q) \cap \{x \in \mathbb{R}^n \mid x_{i_t} \in \{0,1\}\}) \cup conv(P \cap Q \cap \{x \in \mathbb{R}^n \mid x_{i_t} = 1\}))$. Or, simpler: $conv(S_1 \cup S_2) = conv(conv(S_1) \cup conv(S_2))$.
$\to conv(P \cap Q \cap \{x \mid x_{i_t} \in \{0,1\}\}) = conv(P \cap \{x \in \mathbb{R}^n \mid x_k \in \{0,1\} \forall k \in \{i_1,\dots,i_t\}\})$.
Note that $conv(P \cap \{x \mid a^T x = \alpha\} \cap \mathbb{Z}^n) \neq conv(P \cap \mathbb{Z}^n) \cap \{x \mid a^T x = \alpha\}$. However, we can do this in the previous proof because $x_{i_t} = 0$ is a face of P and $x_{i_t} = 1$ is also a face of P.

Example: $P = \{x \mid 2x_1 - x_2 \geq 0, 2x_1 + x_2 \leq 2, 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1\}$. Select $j=1$ and multiply all inequalities by x_1 and $(1-x_1)$: $2x_1^2 - x_1x_2 \geq 0$
$2x_1(1-x_1) - x_2(1-x_1) \geq 0$
$2x_1^2 + x_1x_2 \leq 2x_1$
$2x_1(1-x_1) + x_2(1-x_1) \leq 2(1-x_1)$
$0 \leq x_1^2 \leq x_1$
$0 \leq x_1(1-x_1) \leq 1-x_1$
$0 \leq x_1x_2 \leq x_1$
$0 \leq x_2(1-x_1) \leq 1-x_1$
Now, replace x_1^2 by x_1 and x_1x_2 by y. We obtain: $2x1 - y \geq 0$
$-x_2 + y \geq 0$
$y \leq 0$
$x_2 - y \leq 2 - 2x_1$
$0 \leq x_1 \leq x_1$
$0 \leq 0 \leq 1 - x_1$
$0 \leq y \leq x_1$
$0 \leq x_2 - y \leq 1 - x_1$
As, $y \leq 0$ and $0 \leq y$, $y = 0$, which implies: $x_1 \geq 0$
$x_2 \leq 0$
$x_2 \leq 2 - 2x_1$
$0 \leq x_1 \leq 1$

$0 \leq x_2 \leq 1 - x_1$
As, $x_2 \leq 0$ and $0 \leq x_2$, $x_2 = 0$. Thus, $P_1 = \{(x_1, x_2) \mid 0 \leq x_1 \leq 1, x_2 = 0\} = conv(P \cap \mathbb{Z}^n)$.

Independence Systems and Matroids

Definition - Independece system Let N be a finite set, $N = \{1, \ldots, n\}$. $\mathcal{I}$ is a collection of subsets of N. $(N, \mathcal{I})$ is an independence system if it has the properties:

- $\emptyset \in \mathcal{I}$
- $A \in \mathcal{I}$ and $B \subseteq A \to B \in \mathcal{I}$

Usually, $\mathcal{I}$ is implicitly defined, for example, look at all vertices in a graph that do not form a cycle.

Examples:

1. Linear independence: $A \in \mathbb{Z}^{m \times n}$, $\mathcal{I}$ is a collection of sets in $\{1, \ldots, n\}$ st $\{A_i \mid i \in \mathcal{I}\}$ is linearly independent whenever $i \in \mathcal{I}$.
2. $G = (V, E)$ is given, $I \subseteq E$ is independent if (V, I) does not contain a cycle and $\mathcal{I} = \{I :$ does not contain a cycle$\}$.
3. Knapsack problem: $\{x \in \{0,1\}^n \mid a_1 x_1 + \cdots + a_n x_n \leq b\}$, $a_i \in \mathbb{Z}_+ \setminus \{0\}$ and $b \in \mathbb{Z}_+ \setminus \{0\}$. $N = \{1, \ldots, n\}$. $I \subseteq N$ is independent if $\sum_{i \in I} a_i \leq b$. $\mathcal{I} = \{I \subseteq N \mid \sum_{i \in I} a_i \leq b\}$ is an independent system. This is not true if $A_{\cdot i} \ngeq 0$.

Definition: Let $(N, \mathcal{I})$ be an independent system. Then each member $I \in \mathcal{I}$ is independent. $I \notin \mathcal{I}$ is dependent. Let $S \subseteq N$, $F \subseteq S$, $F \in \mathcal{I}$. F is a basis of S if we can't enlarge it: $F \cup \{i\} \notin \mathcal{I}, \forall i \in S \setminus \{F\}$. A basis of S is not necessarily a basis of N.

The rank of S is $r(S) = \max\{|F| : F \text{ basis of } S\}$.

$I \notin \mathcal{I}$, $I \subseteq N$ is a circuit if $I \setminus \{i\} \in \mathcal{I}, \forall i \in I$.

A independence system $(N, \mathcal{I})$ is a matroid if $\forall S \subseteq N$, every basis of S has the same cardinality.
Example: Consider $\{x \in \{0,1\}^3 \mid x_1 + x_2 + 2x_3 \leq 2\}$.
$S = \{1, 2, 3\} = N$. $\{1, 2\}$ and $\{3\}$ are basis of S. Both are independent and cannot be enlarged, but they do not have the same cardinality, this way, knanpsack problems are not matroids in general.

Definition: Given a finite set N. A function of a set (set function) $f : 2^N \mapsto \mathbb{R}$ is submodular if $\forall S$, $T \subseteq N$, then $f(S) + f(T) \geq f(S \cup T) + f(S \cap T)$. (supermodular $\leq$ instead of $\geq$)

f non decreasing if $f(S) \leq f(T) \forall S \subseteq T$.

Theorem:

1. The set function f is submodular iff $\forall j, k$, $j \neq k$ and $A \subseteq N \setminus \{j, k\}$ then $f(A \cup \{j\}) - f(A) \geq f(A \cup \{j, k\}) - f(A \cup \{k\})$ (diminishing return property)

2. f is submodular and nondecreasing iff $\forall S, T \subseteq N$: $f(T) \leq f(S) + \sum_{j \in T \setminus S}[f(S \cup \{j\}) - f(S)]$.

   $r(S) + r(T \setminus S) \geq |A \cap S| + |A \cap (T \setminus S)| = |A| = r(T)$

$(N, \mathcal{I})$ is a matroid if, for all $F \subseteq N$, every maximal independent set contained in F has the same cardinality $r(F)$.

Theorem: $(N, \mathcal{I})$ is a matroid iff its rank function is submodular.

Proof:

- "$\rightarrow$": $(N, \mathcal{I})$ ind. system then rank function nondecreasing.

  $r(S \cup \{j\}) - r(S) \leq 1$, $r(\emptyset) = 0$. Let $(N, \mathcal{I})$ be a matroid: r satisfies diminishing return property in previous theorem. We want to show that $r(S \cup \{j\}) - r(S) \geq r(S \cup \{j, k\}) - r(S \cup \{k\})$.
  This is satisfied whenever $r(S \cup \{j\}) - r(S) = 1$ or $r(S \cup \{j, k\} - r(S \cup \{k\}) = 0$.
  Suppose $r(S \cup \{j\}) = r(S) = a$. $r(S \cup \{j, k\}) - r(S \cup \{k\}) = 1 \rightarrow r(S \cup \{j, k\}) = a + 1 \rightarrow a = r(S \cup \{k\})$.
  Let Q be a basis of S st $|Q| = a$. $Q \cup \{i\} \notin \mathcal{I}$, $\forall i \in S \setminus \{Q\}$.
  $Q \cup \{j\} \notin \mathcal{I} \rightarrow Q$ basis of $S \cup \{j\}$.
  $Q \cup \{k\} \notin \mathcal{I} \rightarrow Q$ basis of $S \cup \{k\}$.
  Q is a basis of $S \cup \{j, k\}$, which is a contradiction because $|Q| = a$, $r(S \cup \{j, k\}) = a + 1$.

- "$\leftarrow$": assume r is submodular and nondecreasing.

  Suppose $(N, \mathcal{I})$ is not a matroid. There exists $T \subseteq N$ and basis S_1 and S_2 of T st $|S_1| < |S_2|$.
  Check (2) from previous theorem with $T = S_2$ and $S = S_1$:
  $r(S_2) \leq r(S_1) + \sum_{j \in S_2 \setminus S_1}(r(S_1 \cup \{j\}) - r(S_1)) \rightarrow \exists j \in S_2 \setminus S_1$ st $r(S_1 \cup \{j\}) = r(S_1) + 1$
  $S_1 \cup \{j\} \subseteq T$ and $S_1 \cup \{j\} \in \mathcal{I}$, and this is a contradiction because S_1 was basis of T and thus, should have maximum cardinality.

Optimization over matroids

Consider $\max \sum_{j=1}^{n} c_j x_j$. $(N, \mathcal{I})$ independence system, $c_j \in \mathbb{Z}_+$. $\sum_{j \in S} x_j \leq r(S)$, $\forall S \subseteq N$, $x_j \in \{0, 1\} \forall j$ (to find maximum independence set)
Relaxation problem: Let $P_r = \{x \in [0, 1]^n \mid \sum_{j \in S} x_j \leq r(S), \forall S \subseteq N\}$. $F = P_r \cap \mathbb{Z}^n$.

Theorem: Let r be a submodular function $r(S) \in \mathbb{Z}_+$, $\forall S \subseteq N$, r non decreasing, $r(\emptyset) = 0$. $P_r = conv(P_r \cap \mathbb{Z}^n)$, i.e., all the extreme points are integers.

Proof: Consider $\max \sum c_j x_j$ st $x \in P_r$, and its dual is $\min\{\sum_{S \subseteq N} r(S) y_s \mid \sum_{s: j \in S} y_s \geq c_j, y_s \geq 0, \forall S \subseteq N, c \in \mathbb{Z}^n\}$. Wlog: $c_1 \geq c_2 \geq \cdots \geq c_k > 0 \geq c_{k+1} \geq \cdots \geq c_n$ (the values are sorted).
Define $S^0 = \emptyset$, $S^j = \{1, \ldots, j\}$ and

- $x_j = r(S^j) - r(S^{j-1})$ for $j \leq k-1$ and $x_j = 0$, elsewhere.
- $y_s = 0 \forall S \neq S^1, \ldots, S^k$ and $y_{s^k} = c_k$, $y_{s^j} = c_j - c_{j+1} \forall 1 \leq j \leq k-1$.

x, y are integral vectors and x, y are both non negatives vectors.
$\sum_{j \in T} x_j = \sum_{j \in T, j \leq k} (r(S^j) - r(S^{j-1})) \underbrace{\leq}_{\text{definition of submodularity}} \sum_{j \in T, j \leq k} (r(S^j \cap T) - r(S^{j-1} \cap T) \leq \sum_{j=1}^{k} (r(S^j \cap T) - r(S^{j-1} \cap T)) = r(S^k \cap T) - r(\emptyset) \leq r(T)$.
In the dual, $\sum_{s: j \in S} y_s = y_{s^j} + \cdots + y_{s^k} = c_j \; \forall j \leq k$. $\sum_{s: j \in S, j > k} y_s = 0 > c_j$.
primal value equals to dual value: $\sum_{j=1}^{k} c_j (r(S^j) - r(S^{j-1})) = \sum_{j=1}^{k-1} \underbrace{(c_j - c_{j+1})}_{y_{s^j}} r(S^j) + \underbrace{c_k r(S^k)}_{y_{s^k}}$.

The intersection of two matroids

The previous theorem revisited

$(N, \mathcal{I})$ matroid, $r : 2^N \mapsto \mathbb{Z}_+$ submodular, non decreasing. Consider $\max\{\sum_{i \in N} c_i x_i \mid \sum_{i \in T} x_i \leq r(T) \forall T \subseteq N, x_i \geq 0, \; \forall i \in N\}$. This is a integral optimization problem, i.e., all vertices satisfying its constraints are integral.

Consider a graph $G = (V, E)$ and the set of all forests in G, i.e., $T \subseteq E$, (V, T) has no cycles.

$r(T) = |V| - z$, where z is the number of components of (V, T).

Exercise: show that r is submodular exercise

Then, the $(N, \mathcal{I})$ correspondent independence system is a matroid. Consider, $T \subseteq E$, $\sum_{i \in T} x_i \leq r(T)$ (*).
Assume first (V, T) is connected, so (*) is $\sum_{e \in E} x_e \leq n - 1$. $-x_e \leq 0$, $\forall e \in E \setminus T$.
Assume now (V, T) is not connected, let $(V_1, T_1), \ldots, (V_k, T_k)$ be all components of (V, T). Then, (*) is dominated by inequalities: $\sum_{e \in E(V_i)} x_e \leq |V_i| - 1 \forall i$, therefore we can describe the original problem (find a max-weight forest).

We can walk in vertices instead of vertices: $\forall S \subseteq V$: $\sum_{e \in E(S)} x_e \leq |S| - 1, x_e \geq 0$ (subtour elimination constraints).

Properties of rank function $r(\{1, \ldots, j\}) - r(\{1, \ldots, j-1\}) \in \{0, 1\}$, i.e, or it doesn't change or ir increases by one. $r(\emptyset) = 0$.

Interprete the solution combinatorially. Edges are numbered $\{1, \ldots, n\}$. Take the edge with high score, if the rank increases, keep it, if not, skip it. This is the famous greedy algorithm. It only works if we have a matroid.

Greedy-solution

1. Sort edges $c_{e_1} \geq c_{e_2} \geq \cdots \geq c_{e_k} > 0$, $S = \emptyset$.
2. For $j = 1, \ldots, k$:
   if $(V, S \cup \{e_j\})$ does not create a cycle, then, $S = S \cup \{e_j\}$
3. return S

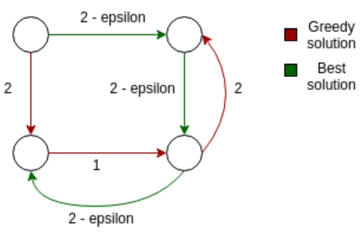
Remark: In lecture 19, we introduced two formulation for the mininum spanning tree problem $G = (V, E)$ connected, $P_{cut} \supseteq P_{SUB}$, $|V| = n$, $|E| = m$. $P_{SUB} = \{x \geq 0 \mid \sum_{e \in E} x_e = n - 1, \sum_{e \in E(S)} x_e \leq |S| - 1 \forall S \subseteq V\}$.

Claim: P_{SUB} **is integral** $Q = \{x \geq 0 \mid \sum_{e \in E(S)} x_e \leq |S| - 1, \forall S \subseteq V\}$ is integral. $Q \cap \{x : \sum_{e \in E} x_e = n - 1\}$ is integral. For all points $x \in Q$ then $\sum_{e \in E} x_e \leq n - 1$. $F = \{x \in Q \mid \sum_{e \in E} x_e = n - 1\}$ is a face of $Q \rightarrow F$ is integral $\rightarrow P_{SUB}$ is integral.

The intersection of two matroids theorem

$(N, \mathcal{I})$ is an independence system. $\mathcal{I} = \mathcal{I}_1 \cap \mathcal{I}_2$, where $(N, \mathcal{I}_i)$ are matroids.

Example: $D = (V, A)$ a digraph. $T \subseteq A$ is a branching if (V, T) contains no cycle and $\forall v \in V$: $|\delta^-(v) \cap T| \leq 1$.



Any branching is an independent set contained in the intersection of the two following matroids:
$\mathcal{I}_1 = \{T \subseteq A \mid |\delta^-(v) \cap T| \leq 1, \forall v \in V\}$ (partition matroid)
$\mathcal{I}_2 = \{T \subseteq A \mid (V, T)$ no cycle $\}$ (graphic matroid)
$\mathcal{I} = \mathcal{I}_\infty \cap \mathcal{I}_\in$.

Greedy algorithm does not work in a branching, i.e, the intersection of two matroids is not a matroid.

Consider optimization problem over intersection of two matroids:
$Z = \max\{\sum_{j=1}^n c_j x_j \mid \underbrace{\sum_{j \in T} x_j \leq min\{r_1(T), r_2(T)\} \forall T \subseteq N, x \geq 0}_{P}\}$, where r_i are rank functions of the two matroids. For functions $f_1(T)$ and $f_2(T)$ as above, $P = conv(F)$

Task: find $x \in P \cap \mathbb{Z}^n$.

Theorem: r_i **are submodular, integer valued,** $r_i(\emptyset) = 0$**,** r_i **non decreasing, then** P **is integral.**

Proof: $\forall c \in \mathbb{Z}^n$, then $z \in \mathbb{Z} \to P$ is integral.

The dual for $\max\limits_{x \in P} c^T x$ **is** $\min\{\sum_{S \subseteq N} r_1(S) y_i(S) + \sum_{S \subseteq N} r_2(S) + y_2(S) \mid \sum_{s: j \in S} y_1(s) + \sum_{s: j \in S} y_2(s) \geq c_j \forall j \in N, y_i(s) \geq 0\}$ (*)

Let (y_1^*, y_2^*) be an optimal dual solution. Let $c_{ji} = \sum_{s: j \in S} y_i^*(s)$ for $i = 1, 2, j \in N$.

Consider: $\min \sum_{s: j \in S} r_i(S) y_i(S)$, for $i = 1, 2$ st $\underbrace{\sum_{s: j \in S} y_i(S) \geq c_{ji} \forall j \in N, y_i(s) \geq 0, \forall S.}_{D_i}$

$\min \sum r_i(S) y_i(S)$, $y_i \in D$ is a dual matroid optimization problem.

From the proof of previou theorem, there exists an alternative dual optimal solution $\bar{y_i}(S), \forall S$ st $\{s : \bar{y_1}(S) > 0\} = \{S_1 \subset S_2 \subset \cdots \subset S_k\}$, $\{s : \bar{y_2}(S) > 0\} = \{U_1 \subset U_2 \subset \cdots \subset U_r\}$.
Thus, there exists an alternative opt. solution (y_1, y_2) for the dual (*) that satisfies:
$Z^{**} = \min \sum_{j=1}^{k} r_1(S_j) y_1(S_j) + \sum_{j=1}^{r} r_2(U_j) y_2(U_j)$ st $By \geq c, y \geq 0$ (**), where B is a matrix with $k + r$ columns. $B_{ij} = 1$ if $i \in S_j$. $j \geq k + 1$, $B_{ij} \in \{0, 1\}$, $B_{ij} = 1$ if $i \in U_{r+1-(j-k)}$ (reverse order).

(**) is a restricted system for (*) by setting many $y_i(S) = 0$. In every row of B, the 1-entries come consecutively. So, this matrix B is TU. c is integral, $r_i(S)$ is also integral $\to y$ optimal for (**) is integral $\to Z^{**}$ is integral $\forall c \in \mathbb{Z}^n$.

Matchings in bipartite graphs

Notation: Let $G = (V, E)$ be a graph. $e \in E$ is viewed as a subset of nodes of cardinality 2.

$deg(v) = |\{v, u\} \in E|$

$deg_F(v) = |\{v, u\} \in F|$, $F \subseteq E$

A path is a sequence of nodes $(v_0, v_1, \ldots, v_t)$ st $\{v_i, v_{i+1}\} \in E$, $\forall i \in 0, \ldots, t-1$

$M \subseteq E$ **is a matching if** $\forall e, e' \in M$, $e \neq e'$, $e \cap e' = \emptyset$ A perfect matching is a matching that covers all vertices, i.e., it has size $\frac{|V|}{2}$.

$F \subseteq E$ **is an edge cover if** $\forall v \in V$, $v \cap e = \emptyset$**, for some** $e \in F$**. Edge cover is the set** F **of edges such that every vertex of** V **is incident to some edge in** F**.** F **covers** V**.**

$W \subseteq V$ **is a node cover if** $\forall e \in E$, $e \cap w = \emptyset$**, for some** $w \in W$**. Node or vertex cover, is the set** W **of vertices that contains at least one endpoint of every edge. This way,** W **covers** E**.**

$S \subseteq V$ **is a stable set if** $e \nsubseteq S$ $\forall e \in E$**. Stable set is a subset of vertices** S **such that no two vertices in** S **are adjacent in** G**.**

Definition - graph parameters

$\alpha(G)$ **is the maximum size of a stable set in** G

$\rho(G) = \min\{|F| \mid F \subseteq E$ **edge cover**$\}$

$\tau(G) = \min\{|W| \mid W \subseteq V$ **node cover**$\}$

$v(G)$ **is the maximum size of a matching in** G

Observations:

1. $S \subseteq V$ stable iff $V \setminus S$ is a node cover
2. $\alpha(G) \leq \rho(G)$
3. $v(G) \leq \tau(G)$

Proof: Consider that G has no isolated node.

1. $S \subseteq V$ stable $\iff \forall e \in E$, $e \nsubseteq S \iff \forall e \in E$, $e \cap (V \setminus S) \neq \emptyset \iff V \setminus S$ is a node cover.
2. Let $S \subseteq V$ be stable and $F \subseteq E$ an edge cover. In particular, F must cover all nodes in S, thus, $\alpha(G) \leq \rho(G)$.
3. Let $M \subseteq E$ be a matching and $W \subseteq V$ is a node cover. In particular, W must cover all edges on M, thus, $v(G) \leq \tau(G)$.

Theorem: For any graph $G = (V, E)$ **without isolated nodes,** $\alpha(G) + \tau(G) = |V| = v(G) + \rho(G)$

Proof: $\alpha(G) + \tau(G) = |V|$ follows from observation.
Let M be a maximum matching, $|M| = v(G)$. M misses $|V| - 2|M|$ nodes. For every "missed" node, pick an edge incident to it. This gives $T \subseteq E$. $M \cup T$ is an edge cover of size $|M| + |V| - 2|M| = |V| - |M| \rightarrow \rho(G) \leq |V| - |M| \rightarrow \rho(G) + \underbrace{|M|}_{=v(G)} \leq |V|$.
Conversely, let $F \subseteq E$ be an edge cover, $|F| = \rho(G)$. Note that if $\{u, v\} \in F$ then $deg_F(u) \leq 1$ or $deg_F(v) \leq 1$. For every $v \in V$ st $deg_F(v) > 1$ remove arbitrarily $deg_F(v) - 1$ many edges from F. This operation leads to a matching of size $|F| - \sum_{v \in V}(deg_F(v) - 1) = |F| + \underbrace{|V|}_{\text{count each node}} - \underbrace{2|F|}_{counteachedgetwice} = -|F| + |V| \rightarrow v(G) \geq |V| - \rho(G) \rightarrow v(G) + \rho(G) \geq |V|$. Thus, $\rho(G) + v(G) = |V|$.

Definition - M-augmenting path $M \subseteq E$ is a matching. A path $(v_0, \ldots, v_t)$ is M-augmenting if:

- t is odd and all nodes $v_i, i = 1, \ldots, t$ are distincts.
- $\{v_1, v_2\}, \ldots, \{v_{t-2}, v_{t-1}\} \in M$.
- v_0 and v_t are nodes missed by M.

Note: if $(v_0, \ldots, v_t)$ is M-augmenting path, then $\{v_0, v_1\}, \{v2, v_3\}, \ldots, \{v_{t-1}, v_t\} \notin M$ and $\bigcup_{i=0}^{t-1}\{v_i, v_{i+1}\} = M \cup M'$, where M' is a matching of size $|M| + 1$.

Remark: Let M be a matching, P an M-augmenting path. Then, $M \Delta E[P]$ gives a larger matching, where $A \Delta B = (A \cup B) \setminus (A \cap B) = (A \setminus B) \cup (B \setminus A)$ is the symmetric difference of the sets A and B.

Theorem: For any graph $G = (V, E)$ a current matching $M \subseteq E$ is either of max size ($|M| = v(G)$) or there exists an M-augmenting path.

Proof: If $|M| = v(G)$ then, no augmenting path exists. Suppose $|M| < v(G)$. Let M' be a matching, $|M'| > |M|$. Consider subgraph $(V, M \cup M')$. G decomposis in connected components: paths, cycles and/or isolated components. As $|M'| > |M|$ there exists a connected component in $(V, M \cup M')$ that is an augmenting path.

Algorithm for computing $v(G)$ when G is bipartite Note: G bipartite means $V = U \cup W$ and $\forall einE$, $e = (u, w)$, where $u \in U$, $w \in W$. It also implies that U is a stable set and W is also a stable set.

1. $M = \emptyset$
2. Find an M-augmenting path and update M as long as possible.

How to find M-augmenting path when G is bipartite? Pick $u \in U$ missed by M. Let $w \in W$ be mised by M. Turn all edges in M from W to U and all edges in $E \setminus M$ from U to W. Test wether there exists a directed path from u to w in the digraph $D = (V, A)$ with arc set A defined before. Using BFS, this task can be accomplished in time $\mathcal{O}(|E|)$.

Theorem: $v(G)$ can be determined, when G is bipartite, in time $\mathcal{O}(|V||E|)$

Konig's Theorem: In a bipartite graph G: $v(G) = \tau(G)$

Proof: From observation, show $v(G) = \tau(G)$. Perform induction on $|V|$
For G contains one edge, the number of vertices is $|V| = 2$, results trivial. **Claim: $\exists u \in V$ that is covered by all max size matchings.**
Consider $G' = G \setminus u$. From hypothesis of induction $v(G') = \tau(G') \to v(G) = v(G') + 1$. Let W' be a node cover in G' of size $\tau(G')$. $W' \cup \{u\}$ is a node cover such that $|W' \cup \{u\}| = \tau(G') + 1 = v(G)$.

Proof of the claim: Suppose not true. For every edge $e = \{u, v\} \in E$, there exists a maximum size matching not covering u or v. So, there exists a maximum matching M covering v but not u and there exists a maximum matchin N covering u but not v. Consider $G' = \{V, M \cup U\}$. The component of G' containing u is a path P. If $|E(P)|$ is odd, P is M-augmenting path, which is a contradiction. Otherwise, if $|E(P)|$ is even, $E(P) \cup \{e\}$ is N-augmenting path, contradiction.

Konig's edge cover Theorem: For any bipartite graph without isolated vertices, $\alpha(G) = \rho(G)$.
Isolated vertice: ?

Matchings in general graphs

Definition: a component of a graph is odd if it has an odd number of vertices. For any graph G, let $\sigma(G)$ be the number of odd components of G. For any graph $G = (V, E)$ and $U \subseteq V$, the graph obtained by deleting all vertices in U and all edges incident with U is denoted by $G\ U$.

For a graph $G = (V, E)$, $v(G) = \min\limits_{U \subseteq V} \frac{1}{2}(|V| + |U| - \sigma(G - U))$.

A graph $G = (V, E)$ has an perfect matching iff $G - U$ has at most $|U|$ odd components for each $U \subseteq V$.

Proof: Follows from previous formula, since G has a perfect matching iff $v(G) \geq \frac{1}{2}|V|$.

Let $G = (V, E)$ be a graph without isolated vertices. Then, $\rho(G) = \max\limits_{U \subseteq V} \frac{|U| + \sigma(G)}{2}$

Finding a maximum-cardinality matching

The max-size algorithm in general graphs: The problem: we want to find an M-augmenting path.

jack Edmonds - 1965: Every time that we find an odd cycle, we shrink the graph by removing the cycle and adding a pseudonode at the cycle place.

Definition of shrink Let X,Y be sets. Shrink Y: $X/Y = \begin{cases} X\text{, if } X \cap Y = \emptyset \\ X \setminus Y \cup \{Y\} \text{where } \{Y\} \text{ is a super node} \end{cases}$.

Let $G = (V, E)$ be a graph and $C \subseteq V$. V/C: replace all nodes in C by a "super node" C. For $e = \{u, v\} \in E$, then $e/C = \begin{cases} \emptyset\text{, if } u \in C \text{ and } v \in C \\ \{u, C\}\text{, if } v \in C, u \notin C \\ \{u, v\}\text{, if } u, v \notin C \end{cases}$
$E/C = \{e/C : e \in E\}$ and $G/C = (V/C, E/C)$

Definition: Let $G = (V, E)$ and $M \subseteq E$. Let $W \subseteq V$ be all nodes missed by M. A walk $(v_0, \ldots, v_t)$ is called M-alternating path if $\{v_i, v_{i+1}\}$ or $\{v_{i-1}, v_i\}$ belongs to M for $i \geq 1$. An M-alternating walk is called an M-blossom if $v_0, \ldots, v_{t-1}$ are all distinct and $v_0 = v_t \in W$, t is odd.

Remark: an M-alternating walk from $v_0 \in W$ to $v_t \in W$ can be found as follows: Let $D = (V, A)$. $A = \{(v, v') : \exists x \in V \text{ st } \{v, x\} \in E \text{ and } \{x, v'\} \in M\}$. A directed path from v_0 to a node adjacent to v_t gives an M-alternating walk in G.

Matching algorithm $\mathcal{A}(M, G, W)$ Parameters: M matching, G graph and W missed nodes.

1. Find a shortest $W - W$-walk, i.e, from a missed node to another missed node in G, P.
2. Check wether P contains M-augmenting path

   If yes, augment the path

   Otherwise, construct M-blossom from P with node set $C \subseteq V$. Let $G = (V/C, E/C)$, $M' = M/C$, W' are missed nodes by M' in G'. Call $\mathcal{A}(M', G', W')$.

If answer for M' is optimal, then M is optimal. If answer is a better matching, then expand it to a matching in G as indicated in the next theorem.

Corollary: $v(G)$ can be computed in time $\mathcal{O}(|V|^2|E|)$ for an undirected graph

Proof: Apply the above algorithm iteratively, starting with $M = \emptyset$ until a maximum-size matching is attained. By using (10?), a shortest M-alternating w-w-walk can be found in time $\mathcal{O}(|E|)$. Moreover, the graph G/C can be constructed in time $\mathcal{O}(|E|)$. Since the recursion has depth at most $|V|$, each application of the algorithm above takes $\mathcal{O}(|V||E|)$ time. Since the number of applications is at most $|V|$, we have the time bound given.

Theorem: Let $C = (v_0, \ldots, v_t)$ be an M-blossom. M attains $v(G)$ $\iff$ $M' = M/C$ attains $v(G/C)$.

Proof:

- "$\rightarrow$"

  Let P be M-augmenting path in G, wlog P does not start in v_0. $P \cap C = \emptyset$ is clear. M-augmenting path in M will be on M'. Otherwise, $P = QR$, where Q is the subpath in P not visiting any node in C, except the last node in Q. Replace last node in Q by supernode C. C is missed by M' and therefore Q/C is M'-augmenting path.

- "$\leftarrow$"

  Let P' be M'-augmenting in $G' = G/C$. If P' does not use "super node" C, then P' is M-augmenting in G. Otherwise, P' ends in the supernode C. Then, $\{u, C\}$ is in $E'(P')$. This corresponds to an edge in E of the form $\{u, v_i\}$.

Assignments

Assignment 01 - 29/09/2017

Topics: Introduction, Linear optimization and extreme points I and II.

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institute for Operations Research
ETH Zurich HG G21-22

Stephan Artmann stephan.artmann@ifor.math.ethz.ch
Christoph Glanzer christoph.glanzer@ifor.math.ethz.ch

IFOR
Institute for Operations Research
Department of Mathematics

Mathematical Optimization — Assignment 1

https://moodle-app2.let.ethz.ch/course/view.php?id=3610

Exercise 1: Knapsack Problem: Continuous and Binary
The **Knapsack Problem** can be formulated as follows

> From n items with given positive values $p_1, p_2, \ldots, p_n$ and positive weights $w_1, w_2, \ldots, w_n$, select items with total weight at most W (a given constant) so as to maximize the total value.

Without loss of generality we assume that the items are sorted such that

$$\frac{p_i}{w_i} \geq \frac{p_{i+1}}{w_{i+1}} \quad i = 1, \ldots, n-1.$$

a) Formulate the Knapsack Problem as a binary integer linear program.

b) Consider the following numerical example with $W = 13$:

j	1	2	3	4	5
p_j	10	80	40	30	22
w_j	1	9	5	4	3

Solve the relaxed linear programming problem, that is, allow all the variables in the problem to be continuous variables in $[0, 1]$, and solve the maximization problem.

c) Suppose now that the variables can only take values in $\{0, 1\}$. Can you solve the maximization problem?

Exercise 2: Polyhedral Cones
A non-empty subset $C \subseteq \mathbb{R}^n$ is called a cone, if (i) $\lambda c \in C$, for all $c \in C$ and for all $\lambda \in \mathbb{R}_{\geq 0}$, and (ii) $c + d \in C$, for all $c, d \in C$.

(a) Let $P = \{x \in \mathbb{R}^n \mid Ax \leq \mathbf{0}\}$ be a polyhedron defined by a matrix $A \in \mathbb{R}^{m \times n}$ and the zero-vector $\mathbf{0} \in \mathbb{R}^m$. Show that P is a cone.

(b) Show that if a non-empty cone $C \subseteq \mathbb{R}^n$ is a polyhedron, then $C = \{x \in \mathbb{R}^n \mid Ax \leq \mathbf{0}\}$ with $A \in \mathbb{R}^{m \times n}$.

Exercise 3: Polyhedral Cone with Extreme Point
A polyhedron of the form $C := \{x \in \mathbb{R}^n : Ax \leq \mathbf{0}\}$ is called a polyhedral cone.

a) Show that the zero vector is the only possible extreme point of C.

b) Assume C has an extreme point. Show that there exists a vector $c \in \mathbb{R}^n$ such that $c^T x > 0$ for all nonzero $x \in C$. Give a construction of such a vector c (depending on A).

Exercise 4: The Minkowski Sum of Convex Sets
For two sets $A, B \subseteq \mathbb{R}^n$, the Minkowski sum is defined as $A + B := \{a + b \in \mathbb{R}^n \mid a \in A, b \in B\}$.

Show that $A + B$ is convex, if both A and B are convex. Is it true that if $A + B$ is convex, then also A and B are convex? Prove the claim, or give a counterexample.

Assignment 02 - 06/10/2017

Topics: Projection of polyhedra, farka's lemma and standard form of polyhedra.

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institute for Operations Research
ETH Zurich HG G21-22

Stephan Artmann stephan.artmann@ifor.math.ethz.ch
Christoph Glanzer christoph.glanzer@ifor.math.ethz.ch

IFOR
Institute for Operations Research
Department of Mathematics

Mathematical Optimization — Assignment 2

https://moodle-app2.let.ethz.ch/course/view.php?id=3610

Exercise 1: Minkowski sum of polyhedra
Let P and Q be polyhedra in $\mathbf{R}^n$. Let $P+Q=\{x+y \mid x\in P, y\in Q\}$ be the Minkowski sum of the two polyhedra.

a) Show that $P+Q$ is a polyhedron.

b) Show that every extreme point of $P+Q$ is the sum of an extreme point of P and an extreme point of Q.

Exercise 2: Facets and extreme points: Canonical examples
a) What are the extreme points of the cube $P=\{x: 0\le x_i\le 1, i=1,\dots,n\}$? How many extreme points are there?

b) Consider the simplex $S=\{x : x_1+\cdots+x_n\le 1, 0\le x_i\le 1, i=1,\dots,n\}$. Prove that $S=\mathrm{conv}(0,e_1,\dots,e_n)$.

Exercise 3: Extreme points of polyhedra
a) Let $P:=\{x\in\mathbf{R}^n \mid Ax=b, x\ge 0\}$ a nonempty polyhedron in standard form. Show that P has an extreme point.

b) We call two vertices v,w of a polyhedron $P=\{x\in\mathbb{R}^n\colon Ax\le b\}$ adjacent if they share an edge, i.e., if there are $n-1$ linearly independent rows $A_{i_1,\cdot},\dots,A_{i_{n-1},\cdot}$ in A such that $\forall k\in\{i_1,\dots,i_{n-1}\}\colon A_{k,\cdot}v=A_{k,\cdot}w=b_k$.

Let $P:=\{x\in\mathbf{R}^n \mid Ax\le b\}$ be a bounded polyhedron, let c be a vector in $\mathbf{R}^n$, and let γ be some scalar. We define
$$Q:=\{x\in P \mid c^Tx=\gamma\}.$$
Assume that Q is nonempty. Show that every extreme point of Q is either an extreme point of P or a convex combination of two adjacent extreme points of P.

Exercise 4: Basic Feasible Solutions
Let $A\in\mathbb{R}^{m\times n}$, $b\in\mathbb{R}^m$, $m\le n$, and suppose that A has full row rank. Consider
$$P=\{x\in\mathbb{R}^n | Ax=b, x\ge 0\}.$$
Prove that $x^*\in P$ is a basic feasible solution for P if and only if there exists $B\subseteq\{1,\dots,n\}$, $|B|=m$ such that

i) the submatrix $A_{\cdot B}\in\mathbb{R}^{m\times m}$ is invertible,

ii) $x^*_i=(A^{-1}_{\cdot B}b)_i\ \forall\, i\in B$,

iii) $x^*_i=0\ \forall\, i\in\{1,\dots,n\}\setminus B$.

Assignment 03 - 13/10/2017

Topics: Duality theorem and representation of polyhedra

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institute for Operations Research
ETH Zurich HG G21-22
Stephan Artmann stephan.artmann@ifor.math.ethz.ch
Christoph Glanzer christoph.glanzer@ifor.math.ethz.ch

IFOR
Institute for Operations Research
Department of Mathematics

Mathematical Optimization — Assignment 3

https://moodle-app2.let.ethz.ch/course/view.php?id=3610

Exercises marked with a (*) may require prior knowledge not yet seen in the lecture, but provide statements that are needed later on in the course.

Exercise 1: Fourier-Motzkin elimination and optimization
Consider the following LP:

$$\begin{array}{llrcrcr} \text{(LP)} & \max & x_1 & + & x_2 & & \\ & s.t. & -\ 4x_1 & - & x_2 & \leq & -8 \\ & & -\ x_1 & + & x_2 & \leq & 3 \\ & & & - & x_2 & \leq & -2 \\ & & 2x_1 & + & x_2 & \leq & 12 \end{array}$$

a) How can the Fourier-Motzkin elimination be modified in order to maximize a linear function over a polyhedron?

b) Apply your idea to the given LP. In particular, derive an optimal solution.

Exercise 2: Farkas Lemma for Standard Form Polyhedra
Prove the following version of the Farkas Lemma: The system $Ax = b$, $x \geq 0$ has a solution if and only if for every y such that $y^T A \geq 0$ it holds that $y^T b \geq 0$.

Exercise 3: Caratheodory's Theorem for Polytopes
Let $P = \text{conv}\{v_1, \cdots v_k\} \subset \mathbb{R}^n$ be a bounded polyhedron (polytope) given as the convex hull of its vertices. Prove that for every point $x \in P$ there exist $n+1$ vertices $v_{i_1}, \cdots, v_{i_{n+1}}$ and nonnegative scalars $\lambda_1, \cdots, \lambda_{n+1}$ with $\sum_{i=1}^{n+1} \lambda_i = 1$ such that

$$x = \sum_{j=1}^{n+1} \lambda_j v_{i_j}.$$

Exercise 4: Iterated Polyhedral Projections
Let $P = \{x \in \mathbb{R}^n \mid Ax \leq b\} \subseteq \mathbb{R}^n$, for $A \in \mathbb{Q}^{m \times n}$, $b \in \mathbb{Q}^m$, be a polyhedron. Show that

$$\text{proj}_{(x_1,\ldots,x_{n-2})}(P) = \text{proj}_{(x_1,\ldots,x_{n-2})}\left(\text{proj}_{(x_1,\ldots,x_{n-1})}(P)\right).$$

Exercise 5: Projection (*)
Let $Q \subset \mathbb{R}^n$ be a non-empty, compact and convex set and let $y \in \mathbb{R}^n$. Consider the projection on Q defined by

$$\mathcal{PO}(y) := \text{argmin}\{||x - y||_2^2 \ : \ x \in Q\}.$$

Show that for all $z \in Q$

$$||z - \mathcal{PO}(y)||_2 \leq ||z - y||_2.$$

Assignment 04 - 20/10/2017

Topics: Simplex algorithm I and II.

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institute for Operations Research
ETH Zurich HG G21-22

Stephan Artmann stephan.artmann@ifor.math.ethz.ch
Christoph Glanzer christoph.glanzer@ifor.math.ethz.ch

IFOR
Institute for Operations Research
Department of Mathematics

Mathematical Optimization — Assignment 4

https://moodle-app2.let.ethz.ch/course/view.php?id=3610

Exercise 1: The Dual program of an LP

(a) Formulate the dual program for the following LP:

$$\begin{array}{ll} \max & x_1 + 2x_2 - x_3 \\ \text{s.t.} & 2x_1 + 3x_2 + 5x_3 \leq 1 \\ & -x_1 + 2x_2 - x_3 \geq 5 \\ & x_1 - 3x_3 = 2 \\ & x_1 \geq 0, x_2 \leq 0, x_3 \in \mathbb{R}. \end{array}$$

(b) Determine the dual program for the following LP:

$$\begin{array}{ll} \max & c^T x + h^T y \\ \text{s.t.} & Ax + By \leq b \\ & Cx + Dy = \gamma \\ & x \in \mathbb{R}^n, y \in \mathbb{R}^p, y \geq 0. \end{array}$$

Exercise 2: Complementary Slackness Conditions

Consider the following (primal) LP:

$$\begin{array}{lccccccccl} \max & 5x_1 & + & 3x_2 & + & x_3 & & & \\ \text{s.t.} & 2x_1 & + & x_2 & + & x_3 & \leq & 6 & \\ & x_1 & + & 2x_2 & + & x_3 & \leq & 7 & \\ & & & & & x_i & \geq & 0 & i = 1, 2, 3 \end{array}$$

Graphically solve the dual of this LP. Then use the complementary slackness conditions to solve the primal problem.

Exercise 3: Infeasibility and LP Duality

(a) Construct an infeasible LP whose dual is feasible. Is it possible to find one such that the dual optimal value is zero?

(b) Construct an infeasible LP whose dual is also infeasible. Explain how you constructed such an example.

Exercise 4: Recession Cone - Extreme Ray with Positive Cost

Let $C := \{x \in \mathbb{R}^n \mid Ax \leq 0\}$ be a pointed cone. Prove the following theorem from the lecture:

Let $c \in \mathbb{R}^n$. Then, $\max\{c^{\mathrm{T}} x \mid x \in C\} = +\infty \Leftrightarrow$ there exists an extreme ray $r \in \mathbb{R}^n$ of C with $c^{\mathrm{T}} r > 0$.

(*) Can you find a solution which does not use the Theorem of Minkowski-Weyl?

Assignment 05 - 27/10/2017

Topics: Representation of Polyhedra and Simplex algorithm III and IV.

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institute for Operations Research
ETH Zurich HG G21-22

Stephan Artmann stephan.artmann@ifor.math.ethz.ch
Christoph Glanzer christoph.glanzer@ifor.math.ethz.ch

Institute for Operations Research
Department of Mathematics

Mathematical Optimization — Assignment 5

https://moodle-app2.let.ethz.ch/course/view.php?id=3610

Exercise 1: Representation of Polyhedra

(a) Consider the polyhedron defined by

$$P := \operatorname{conv}(\{v_1, v_2, v_3\}) + \operatorname{cone}(r) \subset \mathbb{R}^2,$$

where $v_1 = (0,0)^\mathsf{T}, v_2 = (2,1)^\mathsf{T}, v_3 = (0,3)^\mathsf{T}$ and $r = (1,1)^\mathsf{T}$. From the lecture (Theorem of Minkowski-Weyl), we know that P has an *inequality description*. Find A and b such that

$$P = \{x \in \mathbb{R}^2 \mid Ax \leq b\}.$$

(b) Consider the polyhedron defined by

$$P := \{x \in \mathbb{R}^2 \mid x_2 \leq 0,\ x_1 - x_2 \leq 3\}.$$

Find a vertex description of P, namely find $x_1, \dots, x_k \in P$ and $r_1, \dots, r_l \in \mathbb{Z}^n$ such that

$$P = \operatorname{conv}(v_1, \dots, v_k) + \operatorname{cone}(r_1, \dots, r_l).$$

Exercise 2: Complementary Slackness

Consider the following linear program

$$\begin{array}{rrrrrrl}
\max & 7x_1 + & 6x_2 + & 5x_3 - & 2x_4 + & 3x_5 & \\
\text{s.t.} & x_1 + & 3x_2 + & 5x_3 - & 2x_4 + & 2x_5 & \leq 4 \\
 & 4x_1 + & 2x_2 - & 2x_3 + & x_4 + & x_5 & \leq 3 \\
 & 2x_1 + & 4x_2 + & 4x_3 - & 2x_4 + & 5x_5 & \leq 5 \\
 & 3x_1 + & x_2 + & 2x_3 - & x_4 - & 2x_5 & \leq 1 \\
 & & & & & x_j & \geq 0 \quad \text{for } j = 1, \dots, 5
\end{array}$$

of the form $\max\ c^T x$ s.t. $Ax \leq b, x \geq 0$.

(a) Formulate the dual program of the LP.

(b) Consider the primal feasible solution $x^* = (0, \frac{4}{3}, \frac{2}{3}, \frac{5}{3}, 0)^T$. Find a $y^* \in \mathbb{R}^4$ such that

$$\sum_{i=1}^{4} a_{ij} y_i^* = c_j \qquad \text{whenever} \quad x_j^* > 0 \quad \text{for } j = 1, \dots, 5,$$

and

$$y_i^* = 0 \qquad \text{whenever} \quad \sum_{j=1}^{5} a_{ij} x_j^* < b_i \quad \text{for } i = 1, \dots, 4.$$

Is this y^* uniquely determined?

(c) Use the Optimality Condition from complementary slackness theory to prove that the given x^* is **not** an optimal solution for the above linear program.

Exercise 3: Geometry of LP and Exchange Step

Let us consider the following LP problem given in canonical form:

$$
(LP) \qquad \begin{array}{lrrl}
\max & x_1 & +2x_2 & \\
\text{s.t.} & x_1 & +x_2 & \leq 8 \\
 & -x_1 & +x_2 & \leq 2 \\
 & & x_2 & \leq 4 \\
 & x_1 & & \geq 0 \\
 & & x_2 & \geq 0
\end{array}
$$

(a) Draw the feasible set.

(b) Transform (LP) into its corresponding standard form such that the Simplex algorithm is applicable.

(c) For $P_1 = (0,0)$, $P_2 = (0,2)$, $P_3 = (2,4)$, $P_4 = (4,4)$, look at the corresponding point in the standard form P_i', $i = 1,2,3,4$, state which are the basic and non-basic variables and give the corresponding basic solution.

(d) How many exchange steps are necessary to move from point P_1' to P_4'?

(e) Choose to either use the short or long tableau (both are equivalent), construct it corresponding to P_1' and perform the exchange steps to move to P_3':

$$
T_{\text{long}} = \left[\begin{array}{c|cc} -c_B{}^{\mathsf{T}} A_B^{-1} b & 0 \ldots 0 & \bar{c_j},\ j \in N \\ \hline A_B^{-1} b & I & A_B^{-1} A_N \end{array}\right]
$$

$$
T_{\text{short}} = \left[\begin{array}{c|c} -c_B{}^{\mathsf{T}} A_B^{-1} b & \bar{c_j},\ j \in N \\ \hline A_B^{-1} b & A_B^{-1} A_N \end{array}\right].
$$

Have we reached an optimal solution? Why?

Exercise 4: Degeneracy and Cycling of Simplex Method

For this exercise, feel free to use a pivoting tool, such as `http://www.ricoz.net/pivoter.html`.

The goal of this exercise is to show that in case of degeneracy, the finite termination of the Simplex method is not guaranteed by arbitrarily choosing any admissible Simplex pivot.

For the LP

$$\max\{x_1 - 2x_2 + x_3 : (x_1, x_2, x_3) \in \mathbb{R}^3_{\geq 0}\ ;\ 2x_1 - x_2 + x_3 \leq 0\ ;\ 3x_1 + x_2 + x_3 \leq 0\ ;\ -5x_1 + 3x_2 - 2x_3 \leq 0\},$$

show that the following sequence of bases corresponds to a legal execution of the Simplex method:

$$\{4,5,6\} \to \{1,4,6\} \to \{1,3,6\} \to \{1,2,3\} \to \{2,3,5\} \to \{2,4,5\} \to \{4,5,6\},$$

where x_4, x_5 and x_6 are the slack variables corresponding to the constraints $2x_1 - x_2 + x_3 \leq 0$, $3x_1 + x_2 + x_3 \leq 0$ and $-5x_1 + 3x_2 - 2x_3 \leq 0$.

Assignment 06 - 03/11/2017

Topics: Simplex algorithms and Lexicographic Pivoting Rule

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institute for Operations Research
ETH Zurich HG G21-22

Stephan Artmann stephan.artmann@ifor.math.ethz.ch
Christoph Glanzer christoph.glanzer@ifor.math.ethz.ch

IFOR
Institute for Operations Research
Department of Mathematics

Mathematical Optimization — Assignment 6

https://moodle-app2.let.ethz.ch/course/view.php?id=3610

Exercise 1: Modeling linear programs

(a) Consider the LP $\min\{c^\mathsf{T}x \mid x \in \mathbb{R}^n,\ Ax \geq b\}$. How can you find a point in the interior of the feasible set, i.e. in $\{x \in \mathbb{R}^n,\ Ax > b\}$, if one exists?

(b) Assume that X and Y are finite sets of points in $\mathbb{Q}^n$. Write an LP to decide whether there exists a hyperplane $\{x \in \mathbb{R}^n \mid h^\mathsf{T}x = b\}$ with the property that all points in X are lying (strictly) on one side of the hyperplane, while all points in Y are lying (strictly) on the other side. Here, 'strictly' means that we do not want any point to lie directly on the hyperplane.
Hint: The decision whether this is possible is allowed to depend on the resulting objective value.

Exercise 2: Unbounded Simplex Tableau

Consider the following LP:

$$\begin{array}{ll}
\text{minimize} & \begin{pmatrix} 0 \\ -1 \\ -1 \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \\
\text{subject to:} & \begin{pmatrix} 1 & 2 & -3 \\ 2 & -1 & -2 \\ -1 & -1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \leq \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \\
& x_1, x_2, x_3 \geq 0
\end{array}$$

(a) Transform the problem to a LP over a standard-form polyhedron P such that the Simplex algorithm is applicable. Write down a Simplex tableau of your choice corresponding to the transformed problem. Why is it unbounded?

(b) Find a feasible point $x \in P$.

(c) Using the tableau, find $d \in \mathbb{R}^{n+m}$ such that $x + \lambda \cdot d \in P$, for all $\lambda \geq 0$.

Exercise 3: Simplex Phase One

Consider the following LP:

$$\begin{array}{llll}
\text{minimize} & \text{x} & & \\
\text{subject to:} & -x & \leq & -3 \\
& x & \leq & 5 \\
& x & \geq & 0
\end{array}$$

Note that $x = 0$ is not a feasible solution to this LP and we therefore cannot use the slack basis, as we usually did. The goal of this exercise is to find an initial basis for the original problem.

(a) Reformulate the problem above such that we are in a setting in which we can apply the Simplex method.

(b) Make sure that $b \geq 0$ by multiplying the rows by (-1), if necessary.

(c) Formulate the auxiliary LP for the problem above corresponding to "Phase One" of the Simplex method. For a minimization problem of the form $\min\{x \in \mathbb{R}^n \mid Ax = b,\ x \geq 0\}$, where $b \geq 0$, one way to formulate the auxiliary problem is:

$$\min \left\{ \sum_i y_i \;\middle|\; Ax + Iy = b,\ x \geq 0,\ y \geq 0 \right\}.$$

(d) Solve the auxiliary LP using the Simplex method with a pivoting method of your choice. Note that we can now start with a slack basis. Is the original problem feasible? If yes, extract an initial basis from the resulting tableau.

Exercise 4: (*) Lexicographic Pivoting Rule

Show that the Simplex algorithm with the lexicographic pivoting rule always terminates after a finite number of steps.

Hint: For the definition of the lexicographic pivoting rule, see the lecture notes. Start by choosing the extended tableau, where the unit matrix corresponding to the basis variables is situated on the very left:

$$T_{\text{long}} = \left[\begin{array}{c|cc} -c_B{}^\top A_B^{-1} b & 0 \dots 0 & \bar{c}_j,\ j \in N \\ \hline A_B^{-1} b & I & A_B^{-1} A_N \end{array} \right].$$

Let us enumerate the rows of T by $\{0, 1, \dots, m\}$ and the columns of T by $\{0, 1, \dots, n\}$. Observe that initially, for all $i \geq 1$, $T_{i,\cdot} >_{\text{lex}} 0$. Show that this also holds true after a basis exchange. Use this to show that $T_{0,\cdot}^{\text{new}} >_{\text{lex}} T_0$, where $T_{0,\cdot}^{\text{new}}$ is the first row of T after the basis exchange. Conclude that no basis is visited more than once.

Assignment 07 - 10/11/2017

Topics: First question: Interior Point Method and Convex optimization duality I and II.

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institute for Operations Research
ETH Zurich HG G21-22
Stephan Artmann stephan.artmann@ifor.math.ethz.ch
Christoph Glanzer christoph.glanzer@ifor.math.ethz.ch

IFOR
Institute for Operations Research
Department of Mathematics

Mathematical Optimization — Assignment 1

https://moodle-app2.let.ethz.ch/course/view.php?id=3610

Exercise 1: Knapsack Problem: Continuous and Binary
The **Knapsack Problem** can be formulated as follows

> From n items with given positive values $p_1, p_2, \ldots, p_n$ and positive weights $w_1, w_2, \ldots, w_n$, select items with total weight at most W (a given constant) so as to maximize the total value.

Without loss of generality we assume that the items are sorted such that

$$\frac{p_i}{w_i} \geq \frac{p_{i+1}}{w_{i+1}} \quad i = 1, \ldots, n-1.$$

a) Formulate the Knapsack Problem as a binary integer linear program.

b) Consider the following numerical example with $W = 13$:

j	1	2	3	4	5
p_j	10	80	40	30	22
w_j	1	9	5	4	3

Solve the relaxed linear programming problem, that is, allow all the variables in the problem to be continuous variables in $[0, 1]$, and solve the maximization problem.

c) Suppose now that the variables can only take values in $\{0, 1\}$. Can you solve the maximization problem?

Exercise 2: Polyhedral Cones
A non-empty subset $C \subseteq \mathbb{R}^n$ is called a cone, if (i) $\lambda c \in C$, for all $c \in C$ and for all $\lambda \in \mathbb{R}_{\geq 0}$, and (ii) $c + d \in C$, for all $c, d \in C$.

(a) Let $P = \{x \in \mathbb{R}^n \mid Ax \leq \mathbf{0}\}$ be a polyhedron defined by a matrix $A \in \mathbb{R}^{m \times n}$ and the zero-vector $\mathbf{0} \in \mathbb{R}^m$. Show that P is a cone.

(b) Show that if a non-empty cone $C \subseteq \mathbb{R}^n$ is a polyhedron, then $C = \{x \in \mathbb{R}^n \mid Ax \leq \mathbf{0}\}$ with $A \in \mathbb{R}^{m \times n}$.

Exercise 3: Polyhedral Cone with Extreme Point
A polyhedron of the form $C := \{x \in \mathbb{R}^n : Ax \leq \mathbf{0}\}$ is called a polyhedral cone.

a) Show that the zero vector is the only possible extreme point of C.

b) Assume C has an extreme point. Show that there exists a vector $c \in \mathbb{R}^n$ such that $c^T x > 0$ for all nonzero $x \in C$. Give a construction of such a vector c (depending on A).

Exercise 4: The Minkowski Sum of Convex Sets
For two sets $A, B \subseteq \mathbb{R}^n$, the Minkowski sum is defined as $A + B := \{a + b \in \mathbb{R}^n \mid a \in A, b \in B\}$.

Show that $A + B$ is convex, if both A and B are convex. Is it true that if $A + B$ is convex, then also A and B are convex? Prove the claim, or give a counterexample.

Assignment 08 - 17/11/2017

Topics: Total unimodular matrices and applications.

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institute for Operations Research
ETH Zurich HG G21-22
Stephan Artmann stephan.artmann@ifor.math.ethz.ch
Christoph Glanzer christoph.glanzer@ifor.math.ethz.ch

IFOR
Institute for Operations Research
Department of Mathematics

Mathematical Optimization — Assignment 8

https://moodle-app2.let.ethz.ch/course/view.php?id=3610

Exercise 1: Subgradients

Let $f_1(x,y) = -\sqrt{2-x^2-(y-1)^2}, f_2(x,y) = -\sqrt{2-x^2-(y+1)^2}$ be the functions representing the lower half of two balls of radius $\sqrt{2}$ centered in $(0,1,0)$ and $(0,-1,0)$, respectively. We want to consider the function f whose epigraph is the intersection of epif_1 and epif_2, i.e. $f = \max\{f_1, f_2\}$ over $\text{dom}f = \text{dom}f_1 \cap \text{dom}f_2$.

Compute the subdifferential $\partial f(x,y)$ of f for every point of its domain: can you already guess where this set will be empty, where it will be a singleton and where it will contain more than one element? Why?

Exercise 2: Separation of Convex Sets

For this exercise, we will need the following two definitions:

- Two convex sets C and D in $\mathbb{R}^n$ are separated if there is a hyperplane H such that $C \subseteq H^-$ and $C \subseteq H^+$ or vice versa, where H^+ and H^- are the closed halfspaces determined by H.
- Two convex sets C and D in $\mathbb{R}^n$ are strongly separated if there is a $y \in \mathbb{R}^n$, $\alpha < \beta \in \mathbb{R}$ such that $C \subseteq \{x \mid y^\top x \leq \alpha\}$ and $D \subseteq \{x \mid y^\top x \geq \beta\}$, that is, the set $S := \{x \mid \alpha \leq y^\top x \leq \beta\}$ 'separates' C and D.

a) Prove that for two convex sets C and D in $\mathbb{R}^n$ the Minkowski sum $C - D$ is convex and that the following two statements are equivalent:

i) C and D are separated, respectively, strongly separated.

ii) $C - D$ and $\{0\}$ are separated, respectively, strongly separated.

b) Let $C, D \subseteq \mathbb{R}^n$ be convex, C compact, D closed and $C \cap D = \emptyset$. Prove that C and D are strongly separated. You may use the fact that there exist $p \in C, q \in D$ having minimum distance.

Exercise 3: Equality constrained least-squares

Consider the equality constrained least-squares problem

$$\begin{aligned} \min \quad & \|Ax - b\|_2^2 \\ s.t. \quad & Gx = h, \end{aligned}$$

where $A \in \mathbb{R}^{m\times n}$ and where $G \in \mathbb{R}^{p\times n}$.Give the KKT conditions.

Assignment 09 - 24/11/2017

Topics: Algorithms for hard problems and polyhedral representation of discrete optimization problems.

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institute for Operations Research
ETH Zurich HG G21-22
Stephan Artmann stephan.artmann@ifor.math.ethz.ch
Christoph Glanzer christoph.glanzer@ifor.math.ethz.ch

IFOR
Institute for Operations Research
Department of Mathematics

Mathematical Optimization — Assignment 9

https://moodle-app2.let.ethz.ch/course/view.php?id=3610

Exercise 1: Unimodular and Totally Unimodular Matrices

Let $A \in \mathbb{R}^{m \times n}$ and let I be the n-dimensional identity matrix. Prove the following statements:

(a) A is totally unimodular $\Leftrightarrow$ A^{T} is totally unimodular.

(b) A is totally unimodular $\Rightarrow$ $[A \mid I]$ is totally unimodular.

(c) $[A \mid I]$ unimodular $\Rightarrow$ A is totally unimodular.

(d) A is totally unimodular $\Leftrightarrow$ $\begin{bmatrix} A \\ -A \end{bmatrix}$ is totally unimodular.

(e) A is totally unimodular $\Leftrightarrow$ $\begin{bmatrix} A \\ -A \\ I \\ -I \end{bmatrix}$ is totally unimodular.

(f) Assume that $A \in \{0,1\}^{m \times n}$ is a matrix with the *consecutive-ones property*, i.e. for all $i \in \{1,\dots,m\}$, there is $j_1 \le j_2$, $j_1, j_2 \in \{1,\dots,n\}$ such that

$$A_{i,k} = 0,\ k \le j_1 - 1,$$
$$A_{i,k} = 1,\ k \in \{j_1,\dots,j_2\},$$
$$A_{i,k} = 0,\ k \ge j_2 + 1.$$

Prove that A is TU.

(g) Let $G = (V,E)$ be a bipartite graph and $A \in \{0,1,-1\}^{|V| \times |E|}$ be its incidence matrix, i.e. $A_{v,e} = 1 \Leftrightarrow v \in e$ and 0 otherwise. Add an additional row $A_{n+1,\cdot} = (1,1,\dots,1)$ to A. Does the matrix stay TU? Give a proof or counterexample.

(h) Assume that the matrices $[A \mid a] \in \mathbb{Z}^{m_1,n_1+1}$ and $\begin{bmatrix} b^{\mathrm{T}} \\ B \end{bmatrix} \in \mathbb{Z}^{m_2+1 \times n_2}$ are totally unimodular, where a and b are column vectors of appropriate dimension. Show that then, the so-called 2-sum is totally unimodular as well:

$$T := \begin{bmatrix} A & ab^{\mathrm{T}} \\ 0 & B \end{bmatrix}.$$

Note: From statements b) and c) it follows easily that A TU $\Leftrightarrow$ $[A \mid I]$ TU and A TU $\Leftrightarrow$ $[A \mid I]$ unimodular.

Exercise 2: A convex problem in which strong duality fails

Consider the optimization problem

$$\begin{array}{ll} \text{minimize} & e^{-x_1} \\ \text{subject to} & x_1^2/x_2 \le 0 \\ & x \in D \end{array}$$

where $D := \{(x_1, x_2) \in \mathbb{R}^2 | x_2 > 0\}$.

(a) Verify that this is a convex optimization problem and find the optimal value.

(b) Give the convex optimization dual maximization problem and find the optimal dual solution. What is the optimal duality gap?

(c) Why does the duality theorem for convex optimization not apply here?

Assignment 10 - 1/12/2017

Topics: From linear to integer optimization and principles of cutting planes generation.

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institute for Operations Research
ETH Zurich HG G21-22

Stephan Artmann stephan.artmann@ifor.math.ethz.ch
Christoph Glanzer christoph.glanzer@ifor.math.ethz.ch

IFOR
Institute for Operations Research
Department of Mathematics

Mathematical Optimization — Assignment 10

https://moodle-app2.let.ethz.ch/course/view.php?id=3610

Exercise 1: The pigeonhole principle as LP-formulation
The pigeonhole principle states that we cannot place $n+1$ pigeons in n holes, where we only allow one pigeon per hole. This statement can be expressed by the infeasibility of a binary integer optimization problem. Assume throughout this exercise that $n \geq 2$.

Consider the following two formulations of the Pigeonhole problem, where we define

$$x_{ij} = \begin{cases} 1, & \text{if pigeon } i \text{ is placed in hole } j, \\ 0, & \text{otherwise.} \end{cases}$$

$$(P1) \qquad \begin{array}{ll} \sum_{j=1}^{n} x_{ij} = 1, & i = 1, \ldots, n+1, \\ x_{ij} + x_{kj} \leq 1, & j = 1, \ldots, n, \text{ and } i, k = 1, \ldots, n+1, \text{ where } i \neq k, \\ x_{ij} \in \{0,1\}, & i = 1, \ldots, n+1, \text{ and } j = 1, \ldots, n. \end{array}$$

$$(P2) \qquad \begin{array}{ll} \sum_{j=1}^{n} x_{ij} = 1, & i = 1, \ldots, n+1, \\ \sum_{i=1}^{n+1} x_{ij} \leq 1, & j = 1, \ldots, n, \\ x_{ij} \in \{0,1\}, & i = 1, \ldots, n+1, \ j = 1, \ldots, n. \end{array}$$

i) Denote by (P_1^R) and (P_2^R) the LP-relaxation of the polyhedra described by the constraints of $(P1)$ and $(P2)$, respectively. Which of the models above is stronger? In other words, show that either $(P_1^R) \subseteq (P_2^R)$ or $(P_2^R) \subseteq (P_1^R)$.

ii) Show that one of these models has the property that its LP-relaxation is infeasible while the other one isn't.

Exercise 2: Knapsack: Bad approximation of optimal solution
Let $a, c \in \mathbb{Z}^n_{\geq 0}$, $b \in \mathbb{Z}_{\geq 0}$ and assume that $a_i \leq b$, for all $i \in \{1, \ldots, n\}$. Consider the binary knapsack problem

$$\max\{c^{\mathrm{T}} x \mid a^{\mathrm{T}} x \leq b, \ x \in \{0,1\}^n\},$$

and its LP relaxation

$$\max\{c^{\mathrm{T}} x \mid a^{\mathrm{T}} x \leq b, \ x \in [0,1]^n\}.$$

Assume that $\frac{c_1}{a_1} \geq \frac{c_2}{a_2} \geq \ldots \geq \frac{c_n}{a_n}$. We observed in class that the optimal solution for the LP-relaxation x^* is given by

$$x_i^* = \begin{cases} 1, & i \leq i^*, \\ \delta, & i = i^* + 1, \\ 0, & \text{otherwise,} \end{cases}$$

where $i^* \in \{1, \dots, n\}$ is the largest number such that $\sum_{i=1}^{i^*} a_i \leq b$ and $\delta := \frac{b - \sum_{i=1}^{i^*} a_i}{a_{i^*+1}}$.

Define the vectors x^1 and x^2 by

$$x_i^1 = \begin{cases} 1, & i \leq i^* \\ 0, & \text{otherwise,} \end{cases}$$

and

$$x_i^2 = \begin{cases} 1, & i = i^* + 1 \\ 0, & \text{otherwise.} \end{cases}$$

We have seen in class that $\max\{c^\mathrm{T} x^1, c^\mathrm{T} x^2\}$ is an $\frac{1}{2}$-approximation for the binary knapsack problem. Let x^*_{BK} be the optimal solution to the binary knapsack problem.

i) Construct an instance of the binary knapsack problem where the solution $c^\mathrm{T} x^1$ is arbitrarily far away from $c^\mathrm{T} x^*_{BK}$.

ii) Construct an instance of the binary knapsack problem where the solution $c^\mathrm{T} x^2$ is arbitrarily far away from $c^\mathrm{T} x^*_{BK}$.

The term 'arbitrary' means that for a given $N \geq 1$, $c^\mathrm{T} x^*_{BK} - c^\mathrm{T} x^1 \geq N$, resp., $c^\mathrm{T} x^*_{BK} - c^\mathrm{T} x^2 \geq N$.

Exercise 3: Integral polyhedra

Let $P \subset \mathbb{R}^n$ be a non-empty polyhedron without lines, i.e. there exists at least one vertex. Prove the following statement:

P is integral if and only if $\max_{x \in P} c^\mathrm{T} x$ is integral ($\in \mathbb{Z}$) or unbounded for all $c \underline{\in \mathbb{Z}^n}$.

Hint: Recall that a pointed polyhedron P is called integral if and only if every vertex of P is integral. One direction follows easily. For the other direction, choose a vertex $x^* \notin \mathbb{Z}^n$ and $c \in \mathbb{Q}^n$ such that x^* is an optimal vertex solution w.r.t. c (why does such a c exist?). Rescale c such that $c \in \mathbb{Z}^n$. Let $j \in \{1, \dots, n\}$ such that $x_j^* \notin \mathbb{Z}$. Now, consider both c and $\bar{c} := c + (1/a) \cdot e_j$, where e_j is the j^{th} unit vector and $a \in \mathbb{Z}$ is sufficiently large such that x^* is still optimal w.r.t. $\bar{c}$. Try to conclude that there exists an integral cost vector for which the optimal objective value is fractional.

Assignment 11 - 8/12/2017

Topics: The lift and project method and independence systems and matroids.

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institute for Operations Research
ETH Zurich HG G21-22

Stephan Artmann stephan.artmann@ifor.math.ethz.ch
Christoph Glanzer christoph.glanzer@ifor.math.ethz.ch

IFOR
Institute for Operations Research
Department of Mathematics

Mathematical Optimization — Assignment 11

https://moodle-app2.let.ethz.ch/course/view.php?id=3610

Exercise 1: Mixed-integer feasibility with a fixed number of integer variables
The mixed-integer feasibility problem is formulated as follows:

Given $A \in \mathbb{Q}^{m \times d}$, $B \in \mathbb{Q}^{m \times n}$ and $c \in \mathbb{Q}^m$, does there exist $x \in \mathbb{Z}^d$ and $y \in \mathbb{R}^n$ such that $Ax + By \leq c$ is satisfied? If yes, find such a pair of x and y.

Under the assumption that d is a constant, solve this problem by using two oracles: One that can solve LP-feasibility in polynomial-time and one that can solve IP-feasibility in polynomial-time if the amount of variables is constant. You may assume that both oracles return feasible solutions if they exist.

Hint: Consider a projection of the LP-relaxation $\{\binom{x}{y} \in \mathbb{R}^{d+n} \mid Ax + By \leq c\}$, to which you can apply the IP-feasibility oracle. You may assume that the IP-feasibility oracle can be applied to projections of polyhedra, i.e., does not require an explicit inequality description.

Exercise 2: Matching and Perfect Matching Polytope
Let $G = (V, E)$ be an undirected graph. From the lecture, we know that the matching polytope for a bipartite graph is given by $BM(G) := \left\{x \in \mathbb{R}^{|E|} \;\middle|\; x(\delta(v)) \leq 1,\ \forall v \in V,\ x \geq 0\right\}$.

A perfect matching is a matching with the additional property that all vertices are "touched" by the edges in the matching, i.e. a perfect matching is a set $F \subseteq E$ such that $|F \cap \{v\}| = 1$ for all $v \in V$.

For general graphs, it can be shown that an inequality description of the matching polytope is given by

$$\begin{aligned} M(G) := \Big\{x \in \mathbb{R}^{|E|} \;\Big|\; & x(\delta(v)) \leq 1,\ \forall v \in V, \\ & x(E[S]) \leq \frac{|S|-1}{2},\ \forall S \subseteq V,\ |S| \text{ odd}, \\ & x \geq 0\Big\}, \end{aligned}$$

and an inequality description of the perfect matching polytope is given by

$$\begin{aligned} PM(G) := \Big\{x \in \mathbb{R}^{|E|} \;\Big|\; & x(\delta(v)) = 1,\ \forall v \in V, \\ & x(\delta(S)) \geq 1,\ \forall S \subseteq V,\ |S| \geq 1,\ |S| \text{ odd}, \\ & x \geq 0\Big\}. \end{aligned}$$

Let us define

$$
\begin{aligned}
PM'(G) := \big\{x \in \mathbb{R}^{|E|} \;\big|\; & x(\delta(v)) \leq 1,\ \forall v \in V, \\
& x(E[S]) \leq \frac{|S|-1}{2},\ \forall S \subseteq V,\ |S| \text{ odd}, \\
& x(E) = \frac{|V|}{2}, \\
& x \geq 0\big\}.
\end{aligned}
$$

Show that $PM(G) = PM'(G)$, i.e. that both descriptions are equivalent. You may assume that $PM(G)$ is a valid inequality description of the perfect matching polytope for general graphs.

Hint: Distinguish whether $|V|$ is odd or even.

Exercise 3: LP Solution and the Normal Cone
Let $P := \{x \in \mathbb{R}^n \mid Ax \leq b\}$ be a non-empty polyhedron without lines. For a vertex $v \in P$, we define its normal cone to be $\operatorname{cone}\big(\{A_{i,\cdot}^{\mathrm{T}} \mid i \in I\}\big)$, where I is the set of tight inequalities at v.

Prove the following statement:
Let x^* be a vertex where $\max\{c^{\mathrm{T}}x \mid x \in P\}$ is attained, i.e. it exists and is finite. Then, c is contained in the normal cone of x^*.

Assignment 12 - 15/12/2017

Topics: The intersection of two matroids and matching in (bipartite) graphs.

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institute for Operations Research
ETH Zurich HG G21-22

Stephan Artmann stephan.artmann@ifor.math.ethz.ch
Christoph Glanzer christoph.glanzer@ifor.math.ethz.ch

IFOR
Institute for Operations Research
Department of Mathematics

Mathematical Optimization — Assignment 12

https://moodle-app2.let.ethz.ch/course/view.php?id=3610

Exercise 1: Submodular Functions
We are given a finite set $N = \{1, \ldots, n\}$ and a function $f : 2^N \mapsto \mathbb{R}_+$. Prove the following two statements.

(a) The function f is submodular if and only if for all j, k, $j \neq k$ and $S \subseteq N \setminus \{j, k\}$:

$$f(S \cup \{j\}) - f(S) \geq f(S \cup \{j, k\}) - f(S \cup \{k\}).$$

Hint: Telescoping sum.

(b) The function f is submodular and non-decreasing (i.e., for all $S \subseteq T \subseteq N$ we have $f(S) \leq f(T)$) if and only if for all $S, T \subseteq N$:

$$f(T) \leq f(S) + \sum_{j \in T \setminus S} (f(S \cup \{j\}) - f(S)).$$

Hint: Use (a) iteratively.

Exercise 2: Matroid-union
Let $M_1 := (E_1, \mathcal{F}_1)$ and $M_2 := (E_2, \mathcal{F}_2)$ be two matroids with $E_1 \cap E_2 = \emptyset$.
Let $M := (E_1 \cup E_2, \mathcal{F})$ with

$$\mathcal{F} = \{F \subset E_1 \cup E_2 \mid F \cap E_i \in \mathcal{F}_i \;\; \forall\, i \in \{1, 2\}\}$$

Show that M is a matroid.

Hint: Use the following equivalent definition of a matroid: An independence system $(E, \mathcal{F})$ is a matroid if for every $I, J \in \mathcal{F}$ with $|I| < |J|$ there exists $a \in J \setminus I$ such that $I \cup \{a\} \in \mathcal{F}$.

Exercise 3: Hamiltonian Paths and Matroid Intersection
A Hamiltonian Path in a directed graph is a directed path that visits every vertex exactly once. The Hamiltonian path problem is defined as follows: Given a directed graph $G = (V, E)$, decide whether it contains a Hamiltonian path or not. This problem is $\mathcal{NP}$-hard.

Show that the Hamiltonian path problem in directed graphs can be modeled as the problem of finding a maximum-cardinality independent set in the intersection of three matroids defined on the arc set of the graph. In other words, prove that finding a maximum-cardinality independent set in the intersection of three matroids is NP-hard.

Hint: The partition matroid is defined as follows: Given a ground set N, disjoint sets $B_1, \ldots, B_k$ and $d_1, \ldots, d_k \in \mathbb{Z}$ such that $0 \leq d_i \leq |B_i|$ for all i, define $\mathcal{I} := \{I \subseteq N \text{ s.t. } \forall i\colon |I \cap B_i| \leq d_i\}$.

Solutions

Assignment 01

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institute for Operations Research
ETH Zurich HG G21-22

Stephan Artmann stephan.artmann@ifor.math.ethz.ch
Christoph Glanzer christoph.glanzer@student.ethz.ch

IFOR
Institute for Operations Research
Department of Mathematics

Mathematical Optimization — Solution 1

https://moodle-app2.let.ethz.ch/course/view.php?id=3610

Exercise 1: Knapsack Problem and Branch and Bound

a) The integer program (IP) for the Knapsack Problem can be formulated as follows

$$\begin{aligned} \max \quad & \sum_{i=1}^{n} p_i x_i \\ \text{s.t.} \quad & \sum_{i=1}^{n} w_i x_i \leq W \\ & x_i \in \{0,1\} \quad i = 1,\dots,n \end{aligned}$$

The variables x_i, $i = 1, \dots, n$, are interpreted as

$$x_i = \begin{cases} 1 & \text{if item } i \text{ is chosen} \\ 0 & \text{otherwise.} \end{cases}$$

b) The relaxed LP problem is

$$\begin{aligned} \max \quad & \sum_{i=1}^{n} p_i x_i \\ \text{s.t.} \quad & \sum_{i=1}^{n} w_i x_i \leq W \\ & x_i \leq 1 \quad i = 1,\dots,n \\ & x_i \geq 0 \quad i = 1,\dots,n \end{aligned} \tag{1}$$

Trivially, if $\sum_{i=1}^{n} w_i \leq W$, we choose all items, and if $\forall i \in \{1, \dots, n\}\colon w_i > W$, we choose no item. So let us assume from now on that we are in neither of these two cases. Then, the following holds:

Lemma 1
Assume that $p_1/w_1 > p_2/w_2 > \cdots > p_n/w_n$. Denote by $k \in \{1, \dots, m\}$ the index such that $\sum_{i=1}^{k} w_i \leq W$, $\sum_{i=1}^{k+1} w_i > W$. Then an optimal solution to (1) is given by $x^*(W)$, where

$$x_i^*(W) = \begin{cases} 1, & \text{if } i \leq k, \\ \frac{W - \sum_{i=1}^{k} w_i}{w_{k+1}}, & \text{if } i = k+1, \\ 0, & \text{otherwise} \end{cases}.$$

We give a sketch of the proof for those who are interested.

Proof (Sketch). First, observe that since by assumption $(1, \dots, 1)$ is not a solution to (1), there exists an optimal solution y to (1) which fulfills $w^{\mathrm{T}} y = W$. It thus suffices to show that $x^*(W)$ is optimal for

$$
\begin{array}{llcll}
\max & \sum_{i=1}^{n} p_i x_i & & & \\
\text{s.t.} & \sum_{i=1}^{n} w_i x_i & = & W & \\
 & x_i & \leq & 1 & i = 1, \ldots, n \\
 & x_i & \geq & 0 & i = 1, \ldots, n
\end{array} \tag{2}
$$

Let y be optimal for (2). We first show that for all $1 \leq i \leq k : y_i = 1$: Namely, if there was an $l \leq k$ with $y_l < 1$, $w^{\mathrm{T}} y = w^{\mathrm{T}} x^*(W)$ would imply that there is a $o > k$ such that $y_o > 0$. This would imply that there exists $\epsilon > 0$ small enough so that z fulfills $0 \leq z \leq 1$, with z being

$$
z_i = \begin{cases} y_i, & \text{if } i \neq l, o, \\ y_i + \epsilon, & \text{if } i = l, \\ y_i - \frac{w_l}{w_o}\epsilon, & \text{if } i = o \end{cases}.
$$

Note that $w^{\mathrm{T}} z = \sum_{i \neq l,o} w_i y_i + w_l(y_l + \epsilon) + w_o(y_o - w_l \epsilon / w_o) = w^{\mathrm{T}} y = W$ and thus, z is feasible for (2). But then, since $p_l / w_l > p_o / w_o$, $p^{\mathrm{T}} z = p^{\mathrm{T}} y + p_l \epsilon - p_o w_l \epsilon / w_o > p^{\mathrm{T}} y$, contradicting our choice of y.

Thus, both $x^*(W)$ and y are solutions to $\max\{p^{\mathrm{T}} x \colon w^{\mathrm{T}} x = W,\ 0 \leq x \leq 1, x_1 = \cdots = x_k = 1\}$. It is not hard to see that by the choice of our ordering, $y = x^*(W)$. □

By the above Lemma, we can apply the following algorithm:

(i) Sort the variables according to their 'efficiency' $\frac{p_i}{w_i}$.

(ii) Fill the knapsack with items $1, 2, \ldots$ consecutively as long as the total weight does not exceed W.

(iii) If the total weight is still strictly less than W, fill the next unused item with the maximum fraction possible.

The variables are already sorted by their 'efficiency' $\frac{p_i}{w_i}$:

$$
\frac{10}{1} > \frac{80}{9} > \frac{40}{5} > \frac{30}{4} > \frac{22}{3}.
$$

The algorithm of the above lemma yields as an optimal solution $x^*(P) = (1, 1, \frac{3}{5}, 0, 0)^T$ with optimal value $z(P) := 114$.

c) The solution to the linear relaxation provides an upper bound on the optimal value of the IP, hence, $UB := 114$ is an upper bound for our IP. We use the (trivial) solution $(0, 0, 0, 0, 0)^T$ to get the (trivial) lower bound $LB := 0$ for our IP.

Denote by $P_{x_J = d}$ the problem $\max\{p^{\mathrm{T}} x \colon w^{\mathrm{T}} x = W,\ 0 \leq x \leq 1, x_J = d\}$, where $J \subseteq \{1, \ldots, n\}$ and $d \in \{0, 1\}^{|J|}$. We will gradually set certain variables of x to 0 and 1 to find an optimal integral solution.

We first 'branch' (i.e. decide on a variable value) on x_3 (which is an arbitrary choice), cf. Figure 1. Call $(P_{x_{\{3\}} = 1}) = (P_1)$. Note that (P_1) can be reduced to a Knapsack problem, since it can be written as $\max\{10x_1 + 80x_2 + 40 + 30x_4 + 22x_5 \colon x_1 + 9x_2 + 5 + 4x_4 + 5x_3 \leq 13\}$, which can be reformulated as $\max\{10x_1 + 80x_2 + 30x_4 + 22x_5 \colon x_1 + 9x_2 + 4x_4 + 5x_3 \leq 8\}$. We can thus solve it using the algorithm from part b) of this exercise.

An optimal solution of (P_1) is $x(P_1) = (1, \frac{7}{9}, 1, 0, 0)^T$ with $z(P_1) = 112.2$. We now consider subproblems of (P_1) by branching on the variable x_2 (again an arbitrary choice). Denote by (P_{11}) the problem $(P_{x_{\{2,3\}} = (1,1)})$. As $w_2 + w_3 = 14 > W$, (P_{11}) is infeasible and hence, no subproblem of (P_{11}) needs to be considered.

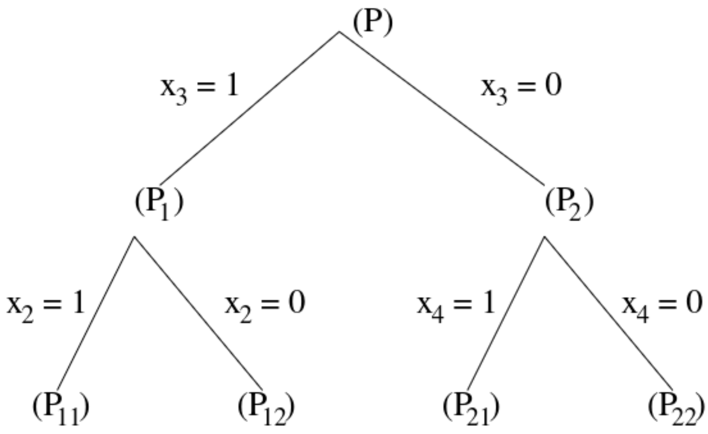


Figure 1: Illustration of our 'branch and bound' approach.

Let (P_{12}) denote the problem $(P_{x_{\{2,3\}}=(0,1)})$. An optimal solution of (P_{12}) (which can again be found by stating the problem as a Knapsack problem) is $x(P_{12}) = (1,0,1,1,1)^T$ with optimal value $z(P_{12}) = 102$. $x(P_{12})$ is integral, and thus we may update our lower bound for the optimal solution: $LB := 102$. Furthermore, since $x(P_{12})$ is integral, no subproblem of P_{12} needs to be considered.

Let (P_2) denote the problem $(P_{x_{\{3\}}=(0)})$. Our relaxed Knapsack algorithm yields the optimal solution $x(P_2) = (1,1,0,\frac{3}{4},0)^T$ with $z(P_2) = 112.5$. We now consider subproblems of P_2 by branching on the variable x_4 (yet again, this is an arbitrary choice), and call $(P_{x_{\{3,4\}}=(1,1)}) =: (P_{21})$ and $(P_{x_{\{3,4\}}=(1,0)}) =: (P_{22})$. The optimal solution of (P_{21}) is $x(P_{21}) = (1,\frac{8}{9},0,1,0)^T$ with $z(P_{21}) = 111.1$. The optimal solution of P_{22} is $x(P_{22}) = (1,1,0,0,1)^T$ with $z(P_{22}) = 112$. $x(P_{22})$ is integral. We update the lower bound: $LB := 112$. As $x(P_{22})$ is integral, no subproblem of P_{22} needs to be considered. Also, no subproblem of P_{21} needs to be considered because $z(P_{21}) = 111.1 < 112 = LB$. As there is no 'open' branch/subproblem left, the optimal solution is $x(P_{22}) = (1,1,0,0,1)^T$ and the optimal objective function value is 112.

Exercise 2: Polyhedral Cones

(a) We check condition (i) and (ii) in the definition of a cone. For (i) let $c \in P$, i.e., $Ac \leq 0$. It follows that $A(\lambda c) = \lambda(Ac) \leq 0$ for all $\lambda \geq 0$. For (ii) let $c, d \in P$, i.e., $Ac \leq 0$ and $Ad \leq 0$. This implies $A(c+d) = Ac + Ad \leq 0$ and thus $c + d \in P$. Therefore, P is a cone.

(b) As C is a polyhedron, there are $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$ with $C = \{x \in \mathbb{R}^n \mid Ax \leq b\}$. We need to prove that $b = 0$. Each nonempty cone contains the origin $\mathbf{0}$ (choose any $x \in C$,then also $0x \in C$). Thus, it holds $A\mathbf{0} = \mathbf{0} \leq b$.

Next, assume there was an $\bar{x} \in P$ with $(A\bar{x})_i > 0$. Then, were would exist a $\bar{\lambda} \geq 0$ with $(A\bar{\lambda}\bar{x})_i = \bar{\lambda}(A\bar{x})_i = b_i$. For any $\lambda > \bar{\lambda}$ the point $\lambda\bar{x} \notin C$ as $(A\lambda\bar{x})_i > b_i$, which would contradict condition (i) in the definition of a cone. We may thus set $b = \mathbf{0}$.

Exercise 3: Polyhedral Cone with Extreme Point

a)Let $x \in C$ be nonzero. Then $3x/2 \in C$, $x/2 \in C$, and $\frac{3x/2+x/2}{2} = x$. Thus x is not an extreme point.

b)Given $C := \{x \in \mathbb{R}^n : Ax \leq \mathbf{0}\}$, set $c^T := -\mathbf{1}^T A$. Since $x = 0$ is an extreme point, the only solution to the equation system $Ax = \mathbf{0}$ is the trivial one, as otherwise it would contain a whole line and having an extreme point is equivalent to not containing a line. In other words, for every non-zero $x \in C$ there is an index i such that $(Ax)_i < 0$ while $(Ax)_j \leq 0$ for all other $j \neq i$. Thus, we have $c^T x = -\mathbf{1}^T Ax > 0$ for every non-zero $x \in C$.

Exercise 4: The Minkowski Sum of Convex Sets
We have to show that $\lambda x + (1-\lambda)y$ is contained in $A+B$ for all $x, y \in A+B$ and all $\lambda \in [0,1]$. Let $x, y \in A+B$. By definition, there are $a_x, a_y \in A$ and $b_x, b_y \in B$ such that $x = a_x + b_x$ and $y = a_y + b_y$. Now consider

$$\lambda x + (1-\lambda)y = \lambda(a_x + b_x) + (1-\lambda)(a_y + b_y) = \big(\ \lambda a_x + (1-\lambda)a_y\ \big) + \big(\ \lambda b_x + (1-\lambda)b_y\ \big).$$

As A and B are convex, it follows that $\lambda a_x + (1-\lambda)a_y \in A$ and $\lambda b_x + (1-\lambda)b_y \in B$. This shows that $\lambda x + (1-\lambda)y \in A+B$.

The opposite direction is not true, however. Consider for example the following sets on real line

$$A = \{0, 1\} \quad B = [0, 1].$$

While A is clearly not convex, we have $A + B = [0, 2]$, which is a convex set.

Assignment 02

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institute for Operations Research
ETH Zurich HG G21-22

Stephan Artmann stephan.artmann@ifor.math.ethz.ch
Christoph Glanzer christoph.glanzer@ifor.math.ethz.ch

IFOR
Institute for Operations Research
Department of Mathematics

Mathematical Optimization — Solution 2

https://moodle-app2.let.ethz.ch/course/view.php?id=3610

Exercise 1: Minkowski sum of polyhedra

We use the following fact: the projection of a polyhedron on a linear space is a polyhedron.

a) Let $P := \{x \in \mathbb{R}^n : Ax \leq b\}$ and $Q := \{x \in \mathbb{R}^n : Cx \leq d\}$. Define $S = \{(x, y, z) \in \mathbb{R}^{3n} : Ax \leq b, Cy \leq d, x + y - z = 0\}$. Then $P + Q$ is the projection of the polyhedron S on its last n components, and is therefore a polyhedron.

b) Let z be an extreme point of $P + Q$. In particular, there exist $x \in P$ and $y \in Q$ such that $x+y = z$. Since z is an extreme point, there exists $c \in \mathbb{R}^n$ s.t. $c^T w < c^T z$ for all $w \in (P+Q) \setminus \{z\}$. Assume there exists $x' \in P$ with $c^T x' \geq c^T x$. Then we would have $c^T(x' + y) \geq c^T(x + y) = c^T z$ and $x' + y \in (P + Q) \setminus \{z\}$, contradicting the fact that z is a vertex. Hence $c^T x' < c^T x$ for all $x' \in P \setminus \{x\}$, so x is an extreme point of P. The same proof can be repeated to show that $y \in Q$ is an extreme point.

Exercise 2: Facets and extreme points: Canonical examples

a) The extreme points of the cube are the points $\{0, 1\}^n$, so there are 2^n extreme points.

b) The points $\{0, e_1, \ldots, e_n\}$ are feasible and basic feasible solutions (though we do not need the latter). Thus $S \supseteq \mathrm{conv}(e_1, \ldots, e_n, 0)$.

To prove equality, choose any $x \in S$. Then $1 \geq \sum_{i=1}^n x_i =: 1-\lambda \geq 0$, and for all $i \in \{1, \ldots, n\}$: $0 \leq x_i \leq 1$. Thus, we can write x as

$$x = \left[\mathcal{I} \mid 0\right] \begin{pmatrix} x \\ \lambda \end{pmatrix},$$

and since $0 \leq \binom{x}{\lambda} \leq 1, \sum_{i=1}^n x_i + \lambda = 1$,

$$\left[\mathcal{I} \mid 0\right] \begin{pmatrix} x \\ \lambda \end{pmatrix} \in \mathrm{conv}(e_1, \ldots, e_n, 0).$$

Therefore, $S = \mathrm{conv}(e_1, \ldots, e_n, 0)$.

Exercise 3: Extreme points of polyhedra

a) $P \subseteq \mathbb{R}^n_{\geq 0}$. Since $\mathbb{R}^n_{\geq 0}$ does not contain a line, the result follows from the theorem covered in the lecture that states that a nonempty polyedron has an extreme point if and only if it does not contain a line.

b) Let x^* be an extreme point of Q. Let $I = I(x^*) := \{i \in \{1, \ldots, m\} : A_i x^* = b_i\}$. x^* is basic feasible for Q, and $c^T x^* = \gamma$, so that we have $\mathrm{rank}(< A_i, c >: i \in I) = n$. We distinguish two cases:

i) if c is linearly dependent with $\{A_i : i \in I\}$, then $\mathrm{rank}(< A_i >: i \in I) = n$ and x^* is an extreme point of P.

ii) Otherwise, $\dim(< A_i >: i \in I) = n-1$, and x^* is contained in the line $L := \{x : A_i x = b_i, i \in I\}$. P is bounded, and thus, so is $P \cap L$. Since $P \cap L$ is a 1-dimensional polyhedron, it is the convex hull of two adjacent extreme points (of $P \cap L$) u and v. Being basic feasible, u fulfills n linearly independent constraints of $P \cap L = \{x \in \mathbb{R}^n\colon Ax \leq b,\ A_I x = b_I\}$, and thus also fulfills n linearly independent constraints of P with equality. Therefore, u is a vertex of P. For the same reason, v is a vertex of P as well. Thus, $x^* \in P \cap L$ is a convex combination of two vertices of P.

Exercise 4: Basic Feasible Solutions
Let us write P as $P = \{x \in \mathbb{R}^n | Cx \leq d\}$, where

$$C = \begin{bmatrix} A \\ -A \\ -\mathcal{I} \end{bmatrix},\ d = \begin{bmatrix} b \\ -b \\ 0 \end{bmatrix}.$$

Let $x^* \in P$. Note that x^* fulfills the first $2m$ constraints with tightness. Let us denote by $\mathcal{I}_n$ the $(n \times n)$-identity matrix.

We have that

x^* is a basic feasible solution for P
$\Leftrightarrow \operatorname{rank}(\{< C_{i\cdot} > | i \in I(x^*)\}) = n$
$\overset{\text{since } \operatorname{rank}(A)=m}{\Leftrightarrow} \exists I \subseteq \{1, \dots, 2m+n\} \colon |I| = n,\ \{1, \dots, m\} \subseteq I,\ \det(C_I) \neq 0,\ C_I x^* = b_I$
$\Leftrightarrow \exists I \subseteq \{1, \dots, 2m+n\} \colon |I| = n,\ \det(C_I) \neq 0,\ C_I x^* = b_I$ and

$$\{C_{i\cdot} | i \in I\} = \underbrace{\{A_{j\cdot} | j = 1, \dots, m\}}_{\text{dimension } m} \cup \{-e_{j_1}, \dots, -e_{j_{n-m}}\}$$

where $\{j_1, \dots, j_{n-m}\} \subseteq \{1, \dots, n\}$ and $j_k \neq j_l$ for all $k \neq l$
$\Leftrightarrow \exists I \subseteq \{1, \dots, 2m+n\}, B \subseteq \{1, \dots, n\} \colon C_I x^* = b_I,\ |B| = m,\ \det(A_{\cdot,B}) \neq 0$ and using row and column permutations, C_I can be transformed into

$$\begin{bmatrix} -\mathcal{I}_{n-m} & 0 \\ A_{\cdot,\{1,\dots,m\}\setminus B} & A_{\cdot,B} \end{bmatrix}$$

$\Leftrightarrow \exists B \subseteq \{1, \dots, m\} \colon |B| = m,\ \det(A_{\cdot,B}) \neq 0,\ x_{\{1,\dots,m\}\setminus B} = 0,\ x_B = (A_{\cdot,B})^{-1} b$ $\square$

Assignment 03

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institute for Operations Research
ETH Zurich HG G21-22

Stephan Artmann stephan.artmann@ifor.math.ethz.ch
Christoph Glanzer christoph.glanzer@ifor.math.ethz.ch

IFOR
Institute for Operations Research
Department of Mathematics

Mathematical Optimization — Solution 3

https://moodle-app2.let.ethz.ch/course/view.php?id=3610

Exercise 1: Fourier-Motzkin elimination and optimization

a) Introduce a new variable z corresponding to the objective function value and add the constraint $z \leq c^{\intercal} x$ to the constraints set. Find $\alpha_1 := \arg\max\{z \in \mathbb{R}\colon z \in \mathrm{proj}_z(P)\}$ - if it does not exist, the original polyhedron is empty, if $\alpha_1 = \infty$, the problem is unbounded.

Otherwise, proceed and find $\binom{\alpha_1}{\alpha_2} \in \{\bar{x} \in \mathbb{R}^2\colon \bar{x} \in \mathrm{proj}_{x_n, z}(P), \bar{x}_1 = \alpha_1\}$, then find $(\alpha_1, \alpha_2, \alpha_3)^{\mathrm{T}} \in \{\bar{x} \in \mathbb{R}^3\colon \bar{x} \in \mathrm{proj}_{x_{n-1}, x_n, z}(P), \bar{x}_{1:2} = \binom{\alpha_1}{\alpha_2}\}$ etc. .

b) We perform Fourier-Motzkin-Elimination as follows:

$$
\begin{array}{lrcrcrcr}
\multicolumn{8}{l}{\textbf{Original System}:} \\
(i) & -\ x_1 & - & x_2 & + & z & \leq & 0 \\
(ii) & -\ 4x_1 & - & x_2 & & & \leq & -8 \\
(iii) & -\ x_1 & + & x_2 & & & \leq & 3 \\
(iv) & & - & x_2 & & & \leq & -2 \\
(v) & 2x_1 & + & x_2 & & & \leq & 12 \\
\multicolumn{8}{l}{\textbf{Projection on } (x_2, z):} \\
2(i) + (v) & & - & x_2 & + & 2z & \leq & 12 \\
(ii) + 2(v) & & & x_2 & & & \leq & 16 \\
2(iii) + (v) & & & 3x_2 & & & \leq & 18 \\
(iv) & & - & x_2 & & & \leq & -2 \\
\multicolumn{8}{l}{\textbf{Removing redundant constraints}:} \\
(i) & & - & x_2 & + & 2z & \leq & 12 \\
(ii) & & & x_2 & & & \leq & 6 \\
(iii) & & - & x_2 & & & \leq & -2 \\
\multicolumn{8}{l}{\textbf{Projection on } z:} \\
(i) + (ii) & & & & & 2z & \leq & 18 \\
(ii) + (iii) & & & 0 & & & \leq & 4
\end{array}
$$

Thus, $z \leq 9$ implies that the maximal objective function value is 9. Setting $z = 9$, the constraint $-x_2 + 2z \leq 12$ leads us to $-x_2 \leq -6$ and together with $x_2 \leq 6$ we obtain $x_2 = 6$. If we plug in $x_2 = 6$ and $z = 9$ into the constraints $-x_1 - x_2 + z \leq 0$ and $2x_1 + x_2 \leq 12$ we obtain $x_1 = 3$. Indeed, $x = \binom{3}{6}$ is a feasible solution with objective function value 9, so we obtained an optimal solution.

Exercise 2: Farkas Lemma for Standard Form Polyhedra

We transform the system $Ax = b,\ x \geq 0$ into the equivalent system

$$
\begin{pmatrix} A \\ -A \\ -I \end{pmatrix} x \leq \begin{pmatrix} b \\ -b \\ 0 \end{pmatrix}
$$

Applying the Farkas lemma, studied in class, the latter system has a solution if and only if the following system has no solution: $y_1, y_2, y_0 \geq 0$,

$$(y_1^T, y_2^T, y_0^T)\begin{pmatrix} A \\ -A \\ -I \end{pmatrix} = 0$$

and

$$(y_1^T, y_2^T, y_0^T)\begin{pmatrix} b \\ -b \\ 0 \end{pmatrix} < 0.$$

The latter system can be rewritten as $(y_1 - y_2)^T A = y_0$ and $(y_1 - y_2)^T b < 0$. It is solvable if and only if the system $z \in \mathbb{R}^m$, $z^T A \geq 0$ and $z^T b < 0$ is solvable (which can be seen by observing that $y_0 \geq 0$ and by replacing $z = y_1 - y_2$). This finishes the proof.

Exercise 3: Caratheodory's Theorem for Polytopes
Let k be the number of vertices $v_1, \dots, v_k$. Without loss of generality, $k \geq n+1$ (so that we can talk of subsets of the vertex set of size $n+1$). Let $x \in \operatorname{conv}(v_1, \dots, v_k)$, then x can be written as $x = \sum_{j=1}^k \lambda_i v_i$, with $\lambda_i \geq 0$ and $\sum_i \lambda_i = 1$. Thus,

$$\begin{pmatrix} x \\ 1 \end{pmatrix} = \sum_i \lambda_i \begin{pmatrix} v_i \\ 1 \end{pmatrix}.$$

Consequently, $\binom{x}{1} \in \operatorname{cone}(\binom{v_1}{1}, \dots, \binom{v_k}{1})$. By Caratheodory's Theorem from the lecture, there are indices $\{i_1, \dots, i_{n+1}\} \subseteq \{1, \dots, k\}$ and $u_1, \dots, u_{n+1} \geq 0$ such that

$$\begin{pmatrix} x \\ 1 \end{pmatrix} = \sum_{j=1}^{n+1} \mu_j \begin{pmatrix} v_{i_j} \\ 1 \end{pmatrix}.$$

Therefore, $x = \sum_{j=1}^{n+1} \mu_j v_{i_j}$, $\mu \geq 0$ and $\sum_{j=1}^{n+1} \mu_j = 1$, which implies that $x \in \operatorname{conv}(v_{i_1}, \dots, v_{i_{n+1}})$.

Exercise 4: Iterated Polyhedral Projections
Call $Q := \operatorname{proj}_{(x_1,\dots,x_{n-2})}(P)$, $S := \operatorname{proj}_{(x_1,\dots,x_{n-2})}\left(\operatorname{proj}_{(x_1,\dots,x_{n-1})}(P)\right)$.
Let $\bar{x} \in \mathbb{R}^{n-2}$.

$$\bar{x} \in Q \Leftrightarrow \exists \alpha, \beta \in \mathbb{R}\colon \begin{pmatrix} \bar{x} \\ \alpha \\ \beta \end{pmatrix} \in P \Leftrightarrow \exists \alpha\colon \begin{pmatrix} \bar{x} \\ \alpha \end{pmatrix} \in \operatorname{proj}_{(x_1,\dots,x_{n-1})}(P) \Leftrightarrow \bar{x} \in S.$$

Exercise 5: Projection
Consider the function $f(x) := ||x-y||_2^2$. In an upcoming lecture you will see that since it is convex and continuously differentiable, we have that $\mathcal{PO}(y) \in Q$ minimizes f over Q if and only if

$$\nabla f(\mathcal{PO}(y))^T (z - \mathcal{PO}(y)) \geq 0 \qquad \text{for all } z \in Q.$$

Therefore

$$2(\mathcal{PO}(y) - y)^T (z - \mathcal{PO}(y)) \geq 0 \qquad \text{for all } z \in Q$$

or

$$(y - \mathcal{PO}(y))^T (z - \mathcal{PO}(y)) \leq 0 \qquad \text{for all } z \in Q.$$

Note that

$$||z - \mathcal{PO}(y)||_2^2 = ||(z-y) + (y - \mathcal{PO}(y))||_2^2 = ||z-y||_2^2 + 2(z-y)^T(y - \mathcal{PO}(y)) + ||y - \mathcal{PO}(y)||_2^2.$$

Thus

$$\begin{aligned}
||z - \mathcal{PO}(y)||_2^2 - ||z-y||_2^2 &= 2(z-y)^T(y - \mathcal{PO}(y)) + ||y - \mathcal{PO}(y)||_2^2 \\
&= 2(z-y)^T(y - \mathcal{PO}(y)) + (y - \mathcal{PO}(y))^T(y - \mathcal{PO}(y)) \\
&= (y - \mathcal{PO}(y))^T(2z - \mathcal{PO}(y) - y) \\
&= 2(y - \mathcal{PO}(y))^T(z - \mathcal{PO}(y)) + (y - \mathcal{PO}(y))^T(\mathcal{PO}(y) - y) \\
&\leq (y - \mathcal{PO}(y))^T(\mathcal{PO}(y) - y) = -||y - \mathcal{PO}(y)||_2^2 \leq 0.
\end{aligned}$$

Assignment 04

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institute for Operations Research
ETH Zurich HG G21-22
Stephan Artmann stephan.artmann@ifor.math.ethz.ch
Christoph Glanzer christoph.glanzer@ifor.math.ethz.ch

IFOR
Institute for Operations Research
Department of Mathematics

Mathematical Optimization — Assignment 4

https://moodle-app2.let.ethz.ch/course/view.php?id=3610

Exercise 1: The Dual Program of an LP

(a) By introducing variables y_1, y_2 and y_3 for the three rows of the primal program, where $y_1 \geq 0$, $y_2 \leq 0$ and $y_3 \in \mathbb{R}$, we can upper-bound the objective function $c^{\mathrm{T}}x$ as follows:

$$\begin{aligned} & x_1 + 2x_2 - x_3 \\ \leq\ & y_1(2x_1 + 3x_2 + 5x_3) \\ & \quad + \ y_2(-x_1 + 2x_2 - x_3) \\ & \quad + \ y_3(x_1 - 3x_3) \\ \leq\ & y_1 + 5y_2 + 2y_3. \end{aligned}$$

As this needs to hold for all x_i ($i \in \{1, 2, 3\}$) separately, we may divide every row by x_i and the third row by $\pm x_3$, leading to

$$\begin{array}{lllllllll} \min & y_1 & + & 5y_2 & + & 2y_3 & & \\ & 2y_1 & - & y_2 & + & y_3 & \geq & 1 \\ & 3y_1 & + & 2y_2 & & & \leq & 2 \\ & 5y_1 & - & y_2 & - & 3y_3 & = & -1 \\ & y_1, y_2 & \geq & 0, & y_3 & \in & \mathbb{R}. & \end{array}$$

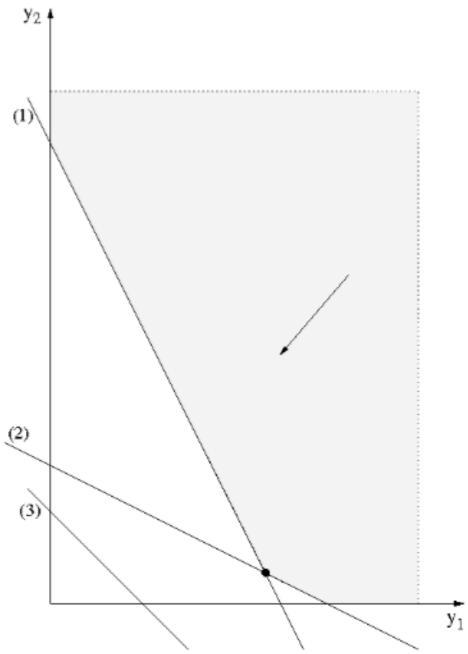
(b) Performing the same operation as in (a) leads to

$$\begin{array}{llllll} \min & y_1^{\mathrm{T}} b & + & y_2^{\mathrm{T}} \gamma & & \\ & A^T y_1 & + & C^T y_2 & = & c \\ & B^T y_1 & + & D^T y_2 & \geq & h \\ & y_1 \geq 0 & & y_2 \text{ free} & & \end{array}$$

Exercise 2: Complementary Slackness Conditions
The dual problem is:

$$\begin{array}{llllllll} \min & 6y_1 & + & 7y_2 & & & & \\ \text{s.t.} & 2y_1 & + & y_2 & \geq & 5 & & (1) \\ & y_1 & + & 2y_2 & \geq & 3 & & (2) \\ & y_1 & + & y_2 & \geq & 1 & & (3) \\ & & & y_i & \geq & 0 & i = 1, 2 & \end{array}$$

The feasible reagion of the dual problem is shown in the following graph:



We graphically derive the dual optimal solution: $(y_1^*, y_2^*) = (\frac{7}{3}, \frac{1}{3})$. Restrictions (1) and (2) are satisfied with equality, whereas restriction (3) is not. Using the complementary slackness conditions, we conclude $x_3^* = 0$. Moreover, as $y_1^* \neq 0, y_2^* \neq 0$, the two primal restrictions have to be satisfied with equality by the optimal solution, i.e. (we already use $x_3^* = 0$):

$$2x_1 + x_2 = 6, \qquad x_1 + 2x_2 = 7$$

From the first equation, we get $x_2 = 6 - 2x_1$. Using the second equation, this yields: $-3x_1 = -5$. Hence the primal optimal solution is $(x_1^*, x_2^*, x_3^*) = (\frac{5}{3}, \frac{8}{3}, 0)$ with a (maximum) objective function value of $16\frac{1}{3}$.

Exercise 3: Infeasibility and LP Duality

(a) LPs of this type are exactly all unbounded LPs. Indeed, suppose a feasible LP max $c^T x$ subject to $Ax \leq b$ and $x \geq \mathbf{0}$ is unbounded, that is, there is z such that $z \geq \mathbf{0}$, $Az \leq \mathbf{0}$ and $c^T z > 0$. Direct application of Farkas' Lemma (Theorem 2.4) yields that the system $-A^T x \leq -c$ and $x \geq \mathbf{0}$ is infeasible. Therefore the dual LP min $b^T y$ subject to $A^T y \geq c$ and $y \geq \mathbf{0}$ is infeasible. Under the assumption that the primal LP is feasible, the opposite direction also holds.

Note that if an LP has optimal solution zero, then, by the strong duality theorem, also its dual is feasible and attains the value zero.

(b) The following LP is an example where both the primal and dual problem are infeasible.

$$\begin{array}{llrcrcrl}
\text{Primal problem:} & \max & x_1 & + & x_2 & & & \\
 & \text{s.t.} & -x_1 & + & x_2 & \leq & -1 & \\
 & & x_1 & - & x_2 & \leq & -1 & \\
 & & & & x_i & \geq & 0 & i = 1,2
\end{array}$$

$$\begin{array}{llrcrcrl}
\text{Dual problem:} & \min & -y_1 & - & y_2 & & & \\
 & \text{s.t.} & -y_1 & + & y_2 & \geq & 1 & \\
 & & y_1 & - & y_2 & \geq & 1 & \\
 & & & & y_i & \geq & 0 & i = 1,2
\end{array}$$

Checking infeasiblity is easy. Adding the two nontrivial constraints in each LP gives $0 \leq -2$ and $0 \geq 2$, respectively.

Exercise 4: Recession Cone - Extreme Ray with Positive Cost

"$\Leftarrow$:" Choose r such that $c^{\mathrm{T}}r > 0$ for the given vector c. Now, $Ar \leq 0$, by assumption. Therefore, for $\lambda \geq 0$, $c^{\mathrm{T}}(\lambda \cdot r) = \lambda \cdot c^{\mathrm{T}}r \to \infty$, $\lambda \to \infty$, while $A(\lambda \cdot r) = \lambda \cdot Ar \leq 0$, i.e. $\lambda \cdot r$ remains feasible for all $\lambda \geq 0$.

"$\Rightarrow$:" We present two solutions, the first of which uses the Theorem of Minkowski-Weyl. As this result is part of the proof of Minkowski-Weyl, we present a second solution which is based on an iterative statement.

- **Solution I**: Using Minkowski-Weyl's Theorem. Assume that for all extreme rays r_i of C, $c^{\mathrm{T}}r_i \leq 0$. Choose any $x \in C$ such that $c^{\mathrm{T}}x > 0$. Such an x must exist as $\max\{c^{\mathrm{T}}x \mid Ax \leq 0\} = +\infty$. Write $x = \sum_{i=1}^{k} \lambda_i r_i$ with $\lambda_i \geq 0$ for all $i \in \{1, \ldots, k\}$. Then, $c^{\mathrm{T}}x = \sum_{i=1}^{k} \lambda_i c^{\mathrm{T}} r_i \leq 0$, a contradiction, as $c^{\mathrm{T}}x > 0$.
- **Solution II**: Using an iterative statement. Choose any $x \in C$ such that $c^{\mathrm{T}}x > 0$. Such an x must exist as $\max\{c^{\mathrm{T}}x \mid Ax \leq 0\} = +\infty$. Denote by I the set of row indices of all tight rows of $Ax \leq 0$ at x. Then, either, $\operatorname{rank}(A_{I,\cdot}) = n - 1$, which implies that x is an extreme ray, or $\operatorname{rank}(A_{I,\cdot}) \leq n - 2$.

  In the second case, proceed as follows: Consider $\{x \in \mathbb{R}^n \mid A_{I,\cdot}x = 0,\ c^{\mathrm{T}}x = 0\}$. By assumption, the kernel of the matrix $\begin{bmatrix} A_{I,\cdot} \\ c^{\mathrm{T}} \end{bmatrix}$ has dimension greater or equal than 1, as $\operatorname{rank}(A_{I,\cdot}) \leq n - 2$ (rank–nullity theorem). Therefore, there exists $l \in \mathbb{R}^n$, $l \neq 0$, such that $A_{I,\cdot}l = 0$ and $c^{\mathrm{T}}l = 0$.

  Let $J := [m] \setminus I$. Consider the partition $J = J_1 \cup J_2$ such that $J_2 \neq \emptyset$ and $A_{J_1,\cdot}l \leq 0$, $A_{J_2,\cdot}l > 0$. We can assume without loss of generality that such a partition exists: As C is pointed, A has full rank, which implies that there is at least one index j such that $A_{j,\cdot}l \neq 0$. If $A_{j,\cdot}l > 0$, we are done. Otherwise, setting $l := -l$ and forming the partition again leads to the desired result.

  Now, define $\lambda^* := \min\left\{-\frac{A_{j,\cdot}x}{A_{j,\cdot}l} \mid j \in J_2\right\}$ (note that $A_{j,\cdot}x < 0$ by assumption). Let index $k \in J_2$ attain the minimum. For $x' := x + \lambda^* l$, it holds that

  - $A_{I,\cdot}x' = A_{I,\cdot}x + \lambda^* A_{I,\cdot}l = A_{I,\cdot}x = 0$,
  - $A_{k,\cdot}x' = A_{k,\cdot}x - A_{k,\cdot}\frac{A_{k,\cdot}x}{A_{k,\cdot}l}l = 0$,
  - $A_{j,\cdot}x' \leq A_{j,\cdot}x - A_{j,\cdot}\frac{A_{j,\cdot}x}{A_{j,\cdot}l}l = 0$, $j \in J_2 \setminus \{k\}$,
  - $A_{j,\cdot}x' = A_{j,\cdot}x - A_{j,\cdot}\frac{A_{k,\cdot}x}{A_{k,\cdot}l}l \leq A_{j,\cdot}x \leq 0$, $j \in J_1$.

  Therefore, $\operatorname{rank}(A_{I(x'),\cdot}) > \operatorname{rank}(A_{I,\cdot})$, while $c^{\mathrm{T}}x' = c^{\mathrm{T}}x > 0$. Iterate this statement until we reach an extreme ray r with $c^{\mathrm{T}}r > 0$.

Assignment 05

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institute for Operations Research
ETH Zurich HG G21-22

Stephan Artmann stephan.artmann@ifor.math.ethz.ch
Christoph Glanzer christoph.glanzer@ifor.math.ethz.ch

Institute for Operations Research
Department of Mathematics

Mathematical Optimization — Solution 5

https://moodle-app2.let.ethz.ch/course/view.php?id=3610

Exercise 1: Representation of Polyhedra

(a) We find the solution graphically. The inequality description is

$$P := \{x \in \mathbb{R}^2 \mid -x_1 \leq 0,\ x_1 - 2x_2 \leq 0,\ x_1 - x_2 \leq 1,\ -x + y \leq 3\}.$$

(b) Again, we find the solution graphically. The vertex description is

$$P = \operatorname{conv}\left(\{(3,0)^\mathsf{T}\}\right) + \operatorname{cone}\left(\{(-1,0)^\mathsf{T}, (-1,-1)^\mathsf{T}\}\right).$$

Exercise 2: Complementary Slackness

(a) We denote the given (primal) linear program as (P). Then, its dual program, (D), is the following.

$$\begin{array}{llllllllllll}
\min & & 4y_1 & + & 3y_2 & + & 5y_3 & + & y_4 & & & \\
\text{s.t.} & & y_1 & + & 4y_2 & + & 2y_3 & + & 3y_4 & \geq & 7 & \\
& & 3y_1 & + & 2y_2 & + & 4y_3 & + & y_4 & \geq & 6 & \\
& & 5y_1 & - & 2y_2 & + & 4y_3 & + & 2y_4 & \geq & 5 & \\
& - & 2y_1 & + & y_2 & - & 2y_3 & - & y_4 & \geq & -2 & \\
& & 2y_1 & + & y_2 & + & 5y_3 & - & 2y_4 & \geq & 3 & \\
& & & & & & & & y_i & \geq & 0 & \text{for } i = 1, \dots, 4
\end{array}$$

(b) For the given $x^* = (0, \frac{4}{3}, \frac{2}{3}, \frac{5}{3}, 0)^T$ we want to find $y^* = (y_1^*, y_2^*, y_3^*, y_4^*)^T$ such that the pair x^*, y^* fulfills the complementary slackness conditions.

Since the second, third, and fourth components of x^* are strictly positive, it follows from the complementary slackness conditions that the second, third, and fourth inequality constraints of (D) must be tight at y^*.

Furthermore, plugging in x^* into the set of constraints defining (P), we see that the third inequality is the only one not satisfied with equality at x^*. By the complementary slackness conditions we infer that $y_3^* = 0$.

Taking both facts together we obtain the following system.

$$\begin{array}{rcrcrcr}
3y_1^* & + & 2y_2^* & + & y_4^* & = & 6 \\
5y_1^* & - & 2y_2^* & + & 2y_4^* & = & 5 \\
-\,2y_1^* & + & y_2^* & - & y_4^* & = & -2
\end{array}$$

The linear system has 3 equations, 3 unknowns, and the 3×3 matrix representing the left-hand side coefficients is nonsingular, thus the linear equation system has a unique solution.

Solving this linear equation system we obtain $y_1^* = y_2^* = y_4^* = 1$, thus the unique dual solution that fulfills the complementary slackness for x^* is $y^* = (1, 1, 0, 1)^T$.

(c) We know that x^* is an optimal solution of (P) if and only if there exists a $y^* = (y_1^*, y_2^*, y_3^*, y_4^*)^T$ such that the pair x^*, y^* fulfills the complementary slackness conditions and y^* is feasible for (D) (i.e., $y^* \geq 0$ and $A^T y^* \geq c$). However, the **unique** y^* such that the pair x^*, y^* fulfills the complementary slackness conditions is not feasible for (D) as it violates the fifth inequality of (D). Therefore, x^* can not be an optimal solution of (P).

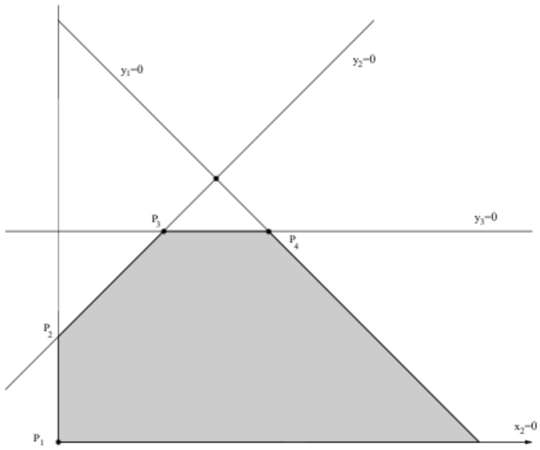


Figure 1: Solution to exercise 3a.

Exercise 3: Geometry of LP and Exchange Step

(a) See Figure 1.

(b)

$$(LPs)\qquad \begin{array}{lrrrrrll} \min & -x_1 & -2x_2 & & & & & \\ \text{s.t.} & x_1 & +x_2 & +y_1 & & & =8 & \\ & -x_1 & +x_2 & & +y_2 & & =2 & \\ & & x_2 & & & +y_3 & =4 & \\ & & x_i & & & & \geq 0 & i=1,2 \\ & & & & & y_i & \geq 0 & i=1,2,3. \end{array}$$

(c)
- $P_1 = (0,0)$:
  basic variables $= \{y_1, y_2, y_3\}$
  non-basic variables $= \{x_1, x_2\}$
  basic solution $= (0,0,8,2,4)$
- $P_2 = (0,2)$:
  basic variables $= \{x_2, y_1, y_3\}$
  non-basic variables $= \{x_1, y_2\}$
  basic solution $= (0,2,6,0,2)$
- $P_3 = (2,4)$:
  basic variables $= \{x_1, x_2, y_1\}$
  non-basic variables $= \{y_2, y_3\}$
  basic solution $= (2,4,2,0,0)$
- $P_4 = (4,4)$:
  basic variables $= \{x_1, x_2, y_2\}$
  non-basic variables $= \{y_1, y_3\}$
  basic solution $= (4,4,0,2,0)$

(d) Since the two sets of basic variables differ by two variables, we need at least two exchange steps. The picture drawn in part (a) shows that this is indeed the right number and it is represented by the path $P_1 \to (8,0) \to P_4$. Clearly, we can also find longer sequences of exchange steps that start at P_1 and end at P_3.

Remark: In general, the number of basic variables in which two vertices are different determines a lower bound on the number of exchange steps needed to move from one vertex to the other.

(e) Using the short tableau and the route via $P_1 \to P_2 \to P_3 \to P_4$, the initial tableau is (using the slack basis):

		x_1	x_2
	0	-1	-2
x_3	8	1	1
x_4	2	-1	1
x_5	4	0	1

Now, when moving to P_4, the short tableaus are:

		x_1	x_4
	4	-3	2
x_3	6	2	-1
x_2	2	-1	1
x_5	2	1	-1

$\rightarrow$

		x_5	x_4
	10	3	-1
x_3	2	-2	1
x_2	4	1	0
x_1	2	1	-1

$\rightarrow$

		x_5	x_3
	12	1	1
x_4	2	-2	1
x_2	4	1	0
x_1	4	-1	1

Thus, the optimal solution is given by -12 and the solution is optimal because the reduced cost vector is ≥ 0. The resulting point corresponds to P_4, which can be verified by setting the non-basic variables to zero.

Remark: This is not the only possible sequence of exchange steps. For instance, we could have moved to $(8, 0)$ and then to P_4, which would have been faster.

Exercise 4: Degeneracy and Cycling of Simplex Method

We transform our given polyhedron to a standard form polyhedron, i.e.

$$\begin{aligned} \text{minimize:} \quad & -x_1 + 2x_2 - x_3, \\ \text{subject to:} \quad & 2x_1 - x_2 + x_3 + x_4 = 0 \\ & 3x_1 + x_2 + x_3 + x_5 = 0 \\ & -5x_1 + 3x_2 - 2x_3 + x_6 = 0 \\ & x_1, x_2, x_3, x_4, x_5 \geq 0. \end{aligned}$$

Therefore, the initial tableau, using the slack basis, is

		x_1	x_2	x_3
	0	-1	2	-1
x_4	0	2	-1	1
x_5	0	3	1	1
x_6	0	-5	3	-2

Now, we perform the exchange steps.

		x_1	x_2	x_3
	0	-1	2	-1
x_4	0	2	-1	1
x_5	0	3	1	1
x_6	0	-5	3	-2

$\rightarrow$

		x_5	x_2	x_3
	0	0.333	2.333	-0.666
x_4	0	-0.666	-1.666	0.333
x_1	0	0.333	0.333	0.333
x_6	0	1.666	4.666	-0.333

$\rightarrow$

		x_5	x_2	x_3
	0	-1	-1	2
x_3	0	-2	-5	3
x_1	0	1	2	-1
x_6	0	1	3	1

$\rightarrow$

		x_5	x_6	x_4
	0	-0.666	0.333	2.333
x_3	0	-0.333	1.666	4.666
x_1	0	0.333	-0.666	-1.666
x_2	0	0.333	0.333	0.333

$\rightarrow$

		x_1	x_6	x_4
	0	2	-1	-1
x_3	0	1	1	3
x_5	0	3	-2	-5
x_2	0	-1	1	2

$\rightarrow$

		x_1	x_6	x_3
	0	2.333	-0.666	0.333
x_4	0	0.333	0.333	0.333
x_5	0	4.666	-0.333	1.666
x_2	0	-1.666	0.333	-0.666

$\rightarrow$

		x_1	x_2	x_3
	0	-1	2	-1
x_4	0	2	-1	1
x_5	0	3	1	1
x_6	0	-5	3	-2

Thus, we have arrived at the original slack basis and the Simplex algorithm does not terminate with this choice of basis changes.

Assignment 06

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institute for Operations Research
ETH Zurich HG G21-22

Stephan Artmann stephan.artmann@ifor.math.ethz.ch
Christoph Glanzer christoph.glanzer@ifor.math.ethz.ch

Institute for Operations Research
Department of Mathematics

Mathematical Optimization — Solution 6

https://moodle-app2.let.ethz.ch/course/view.php?id=3610

Exercise 1: Modeling linear programs

(a) Consider the following LP:

$$\max \left\{ t \;\middle|\; \begin{array}{c} A_{i,\cdot}x \geq b_i + t,\ i = 1, \ldots, n \\ x \in \mathbb{R}^n,\ t \in \mathbb{R}_+ \end{array} \right\}.$$

If $t > 0$ in an optimal solution, then $x^* \in \operatorname{int}(P)$ exists.

(b) Consider the following LP:

$$\begin{array}{llcll} \text{maximize} & \text{t} & & & \\ \text{subject to:} & h^\mathsf{T} x & \leq & b - t & \forall x \in X, \\ & h^\mathsf{T} y & \geq & b + t & \forall y \in Y, \end{array}$$

where $h \in \mathbb{R}^n$, $b \in \mathbb{R}$, $t \geq 0$. If $t > 0$ in an optimal solution, such a hyperplane is given by $\{x \in \mathbb{R}^n : h^\mathrm{T} x = b\}$.

Exercise 2: Unbounded Simplex Tableau

(a) In standard form, the problem is given by

$$\text{minimize} \quad \begin{pmatrix} 0 \\ -1 \\ -1 \\ 0 \\ 0 \\ 0 \end{pmatrix}^\mathrm{T} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix}$$

$$\text{subject to:} \quad \begin{pmatrix} 1 & 2 & -3 & 1 & 0 & 0 \\ 2 & -1 & -2 & 0 & 1 & 0 \\ -1 & -1 & -1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

$$x_1, x_2, x_3, x_4, x_5, x_6 \geq 0$$

The corresponding tableau is:

		x_1	x_2	x_3
	0	0	-1	-1
x_4	0	1	2	-3
x_5	1	2	-1	-2
x_6	0	-1	-1	-1

.

We observe that $\bar{c}_3 < 0$, while the corresponding column admits only non-positive entries. Therefore, the problem is unbounded.

(b) As seen in the lecture, the Simplex algorithm maintains a feasible primal solution when switching bases. Thus, a feasible point is given by the solution corresponding to the current basis, i.e.,

$$x_i = \begin{cases} \left(A_B{}^{-1}b\right)_i & \text{if } i \in B, \\ 0 & \text{else.} \end{cases}$$

Therefore, a feasible solution is given by $x := \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}^{\mathrm{T}}$.

(c) In the lecture, we have seen that in this case, we may consider the feasible direction corresponding to x_3, given by $d := \begin{pmatrix} -A_B^{-1}A_{\cdot,3} \\ e_3 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 3 & 2 & 1 \end{pmatrix}^{\mathrm{T}}$. Indeed, one can check that $c^{\mathrm{T}}d = -1 < 0$, while $Ad = 0$.

Exercise 3: Simplex Phase One

(a) Using slack variables, we can reformulate the problem as:

$$\begin{array}{llcccl} \text{minimize} & \text{x} & & & & \\ \text{subject to:} & -x & + & y_1 & = & -3 \\ & x & + & y_2 & = & 5 \\ & x, y_1, y_2 & & & \geq & 0 \end{array} .$$

(b) The result is:

$$\begin{array}{llccccl} \text{minimize} & \text{x} & & & & & \\ \text{subject to:} & x & - & y_1 & & & = 3 \\ & x & & & + & y_2 & = 5 \\ & x, y_1, y_2 & & & & & \geq 0 \end{array} .$$

(c) One possibility to formulate the auxiliary LP is:

$$\min \left\{ z_1 + z_2 \;\middle|\; \begin{array}{c} x - y_1 + z_1 = 3 \\ x + y_2 + z_2 = 5 \\ x, y_1, y_2, z_1, z_2 \geq 0 \end{array} \right\}.$$

(d) Using that $c_B = (1,1)^\top$, $c_N = (0,0,0)^\top$, $A_B = I$ and $A_N = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$, the corresponding tableau (short) is:

		x	y_1	y_2
	-8	-2	1	-1
z_1	3	1	-1	0
z_2	5	1	0	1

Here, we used that $\overline{c_j} = c_j - c_B^\top A_B^{-1} A_{\cdot,j} = -c_B^\top A_{\cdot,j} \Rightarrow \overline{c} = (-2, 1, -1)^\top$.

By performing the Simplex algorithm, we arrive at the following tableau:

		z_1	y_1	z_2
	0	1	0	1
x	3	1	-1	0
y_2	2	-1	1	1

As the objective value is zero, the original problem is feasible. A feasible basis of the original problem is therefore given by $\{x, y_2\}$.

Exercise 4: Lexicographic Pivoting
We will use the following observations:

i) $u <_{\text{lex}} v \Leftrightarrow u - v <_{\text{lex}} 0$,

ii) For any $v \in \mathbb{R}^n$, $u >_{\text{lex}} 0$ and $\alpha > 0$, it holds that $v + \alpha \cdot u >_{\text{lex}} v$.

As written in the hint, we choose the long tableau where the unit matrix is located to the very left.

$$T_{\text{long}} = \left[\begin{array}{c|cc} -c_B{}^{\mathsf{T}} A_B^{-1} b & 0 \ldots 0 & \bar{c_j},\ j \in N \\ \hline A_B^{-1} b & I & A_B^{-1} A_N \end{array}\right].$$

Let us start by showing that a basis exchange maintains the property that $T_{i,\cdot} >_{\text{lex}} 0$. This clearly holds true for the initial tableau as $A_B^{-1} b \geq 0$ (x is feasible) and due to the unit matrix. Let us change the basis from B to $B \cup \{j\} \setminus \{k\}$. Call the tableau before the basis change T and the new tableau T^{new}. By assumption, $\bar{c}_j < 0$, $\bar{A}_{k,j} > 0$ and by our choice of k, $\frac{T_{k,\cdot}}{\bar{A}_{k,j}} <_{\text{lex}} \frac{T_{i,\cdot}}{\bar{A}_{i,j}}$ for all $i \geq 1, i \neq k$ such that $\bar{A}_{i,j} > 0$.

It follows that,

- $T^{\text{new}}_{k,\cdot} = \frac{1}{\bar{A}_{k,j}} T_{k,\cdot} >_{\text{lex}} 0$ as $\bar{A}_{k,j} > 0$,
- $T^{\text{new}}_{i,\cdot} = T_{i,\cdot} - \frac{\bar{A}_{i,j}}{\bar{A}_{k,j}} T_{k,\cdot}$, $\forall i \geq 0,\ i \neq k$. Let us distinguish between two cases:
  - If $\bar{A}_{i,j} \leq 0$, $T^{\text{new}}_{i,\cdot} >_{\text{lex}} 0$ as $T_{i,\cdot} >_{\text{lex}} 0$, where we use that $T_{k,\cdot} >_{\text{lex}} 0$.
  - For $\bar{A}_{i,j} > 0$, it holds that: $\frac{T_{i,\cdot}}{\bar{A}_{i,j}} >_{\text{lex}} \frac{T_{k,\cdot}}{\bar{A}_{k,j}} \Leftrightarrow T_{i,\cdot} >_{\text{lex}} \frac{\bar{A}_{i,j}}{\bar{A}_{k,j}} T_{k,\cdot} \Leftrightarrow T_{i,\cdot} - \frac{\bar{A}_{i,j}}{\bar{A}_{k,j}} T_{k,\cdot} >_{\text{lex}} 0 \Leftrightarrow T^{\text{new}}_{i,\cdot} >_{\text{lex}} 0$.

This proves that the desired property still holds true after the basis exchange step.

We will now show that $T^{\text{new}}_{0,\cdot} >_{\text{lex}} T_{0,\cdot}$. As $\bar{c}_j < 0$,

$$T^{\text{new}}_{0,\cdot} = T_{0,\cdot} - \frac{\bar{c}_j}{\bar{A}_{k,j}} T_{k,\cdot} >_{\text{lex}} T_{0,\cdot}.$$

It follows that in every basis exchange step, the topmost row of the tableau will increase strictly (lexicographically). We can therefore never visit the same basis twice, as otherwise, we would end up with the same topmost row of T, a contradiction.

Assignment 07

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institute for Operations Research
ETH Zurich HG G21-22

Stephan Artmann stephan.artmann@ifor.math.ethz.ch
Christoph Glanzer christoph.glanzer@ifor.math.ethz.ch

Institute for Operations Research
Department of Mathematics

Mathematical Optimization — Solution 7

https://moodle-app2.let.ethz.ch/course/view.php?id=3610

Exercise 1: Central path

a) First we reformulate the problem to be in the form seen in the lecture:

$$\begin{array}{rll} \min & y_1 + 2y_2 & \\ \text{s.t.} & y_1 + y_2 & \geq 4 \\ & y_2 & \geq 0 \end{array}$$

Then we identify:

$$A^T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \qquad c = \begin{pmatrix} 4 \\ 0 \end{pmatrix} \qquad b = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

Define slack variables s_1, s_2:

$$\begin{array}{rrl} \max & -y_1 - 2y_2 & \\ \text{s.t.} & y_1 + y_2 - s_1 & = 4 \\ & y_2 - s_2 & = 0 \\ & s_1,\ s_2 & \geq 0 \end{array} \quad ,$$

which yields:

$$\begin{aligned} s_1 &= y_1 + y_2 - 4 \\ s_2 &= y_2 \end{aligned}$$

b) The logarithmic barrier function for this problem is given by:

$$\begin{aligned} \phi(y) &= -\sum_{j=1}^{n} \log(s_j) \\ &= -\log(y_1 + y_2 - 4) - \log(y_2). \end{aligned}$$

c) Since the central path $y(\mu)$ is the minimizer of $b^\top y + \frac{1}{\mu}\phi(y)$, we compute the gradient of $\phi(y)$:

$$\begin{aligned} \frac{\partial}{\partial y_1}\phi(y) &= -\frac{1}{y_1 + y_2 - 4} \\ \frac{\partial}{\partial y_2}\phi(y) &= -\frac{1}{y_1 + y_2 - 4} - \frac{1}{y_2} \end{aligned}$$

or directly:

$$\begin{aligned} \nabla\phi(y) &= -A \begin{pmatrix} \frac{1}{s_1} \\ \frac{1}{s_2} \end{pmatrix} \\ &= \begin{pmatrix} -\frac{1}{y_1+y_2-4} \\ -\frac{1}{y_1+y_2-4} - \frac{1}{y_2} \end{pmatrix} \end{aligned}$$

Thus,

$$\begin{aligned}
\nabla\left\{b^\top y + \frac{1}{\mu}\phi(y)\right\} &= b - \frac{1}{\mu}A\begin{pmatrix} \frac{1}{s_1} \\ \frac{1}{s_2} \end{pmatrix} \\
&= \begin{pmatrix} 1 - \frac{1}{\mu(y_1+y_2-4)} \\ \\ 2 - \frac{1}{\mu(y_1+y_2-4)} - \frac{1}{\mu y_2} \end{pmatrix}
\end{aligned}$$

To obtain the minimum of $\phi(y)$ we set the gradient to zero and solve for y_1, y_2. This yields:

$$\begin{aligned}
1 =& \frac{1}{\mu(y_1+y_2-4)} \\
\mu(y_1+y_2-4) =& 1 \\
y_1+y_2-4 =& \frac{1}{\mu}
\end{aligned}$$

as well as

$$\begin{aligned}
\frac{1}{\mu}\left(\frac{1}{y_1+y_2-4} + \frac{1}{y_2}\right) =& 2 \\
\left(\frac{1}{y_1+y_2-4} + \frac{1}{y_2}\right) =& 2\mu \\
\left(-\mu + \frac{1}{y_2}\right) =& 2\mu \\
-y_2\mu + 1 =& 2y_2\mu \\
1 =& y_2\mu \\
y_2 =& \frac{1}{\mu}
\end{aligned}$$

leading to

$$y_1 + \frac{1}{\mu} - 4 = \frac{1}{\mu}$$

and we arrive at

$$y_1 = 4 \qquad y_2 = \frac{1}{\mu}.$$

Hence the analytic path $y(\mu)$ is given by:

$$y(\mu) = \begin{pmatrix} 4 \\ \frac{1}{\mu} \end{pmatrix}$$

Letting $\mu \to \infty$ we obtain an optimal solution as $\binom{4}{0}$.

Exercise 2: Strict vs. Strong Convexity

i) $f : [0,1] \to \mathbb{R}$ defined by $f(0) = f(1) = 1$ and $f(x) = 0$ for $x \in (0,1)$ is convex, but not continuous.

ii) $f(x) = |x|$ is convex, but not strictly convex. In fact, note that $f(\lambda x) = |\lambda| f(x)$, so if $y = 0$ then $f(\lambda x + (1-\lambda)y) = f(\lambda x) \not< \lambda f(x) = \lambda f(x) + (1-\lambda)f(y)$.

iii) $f(x) = e^x$ is strictly convex, but not strongly convex. In fact, $f''(x) = e^x > 0\ \forall x \in \mathrm{dom}(f)$, so it is strictly convex. It is not strongly convex because its second derivative can be arbitrarily close to zero.

iv) $f(x) = x^2$ is strongly convex with modulus 2 because $f''(x) = 2 \ \forall x \in \mathrm{dom}(f)$.

Exercise 3: Convergence of the Newton method
Note that for $k \geq 1$, $\|\nabla f(x_t)\|_2^k \leq \left(\frac{L}{2\sigma^2}\right)^k \|\nabla f(x_{t-1})\|_2^{2k}$. Thus

$$\frac{L}{2\sigma^2}\|\nabla f(x_T)\|_2 \leq \frac{L}{2\sigma^2}\left(\frac{L}{2\sigma^2}\|\nabla f(x_{T-1})\|_2^2\right) \leq \frac{L}{2\sigma^2}\frac{L}{2\sigma^2}\left(\frac{L}{2\sigma^2}\frac{L}{2\sigma^2}\|\nabla f(x_{T-2})\|_2^4\right) \leq \cdots \leq \left(\frac{L}{2\sigma^2}\right)^{2^T} \underbrace{\|\nabla f(x_0)\|}_{\leq \frac{\sigma^2}{L}}{}^{2^T},$$

which in turn fulfills

$$\left(\frac{L}{2\sigma^2}\right)^{2^T} \|\nabla f(x_0)\|_2^{2^T} \leq \left(\frac{1}{2}\right)^{2^T}.$$

Exercise 4: 'Invariance' of the Newton step under linear transformations
By applying the chain rule, we obtain

$$\begin{aligned} \nabla \tilde{f}(x) &= A^T \nabla f(Ax) \\ \nabla^2 \tilde{f}(x) &= A^T \nabla^2 f(Ax) A \end{aligned}$$

Hence we get

$$\begin{aligned} x^{k+1} &\overset{(\mathrm{def})}{=} x^k - (\nabla^2 \tilde{f}(x^k))^{-1} \cdot \nabla \tilde{f}(x^k) \\ &= x^k - A^{-1}(\nabla^2 f(Ax))^{-1}(A^T)^{-1} \cdot A^T \nabla f(Ax) \\ &= x^k - A^{-1}(\nabla^2 f(Ax))^{-1} \nabla f(Ax) \end{aligned}$$

and

$$\begin{aligned} Ax^{k+1} &= Ax^k - AA^{-1}(\nabla^2 f(Ax))^{-1}\nabla f(Ax) \\ &= Ax^k - (\nabla^2 f(Ax^k))^{-1}\nabla f(Ax^k) \\ &\overset{(Ax^k = y^k)}{=} y^k - (\nabla^2 f(y^k))^{-1}\nabla f(y^k) \\ &\overset{(\mathrm{def})}{=} y^{k+1} \end{aligned}$$

Assignment 08

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institute for Operations Research
ETH Zurich HG G21-22

Kevin Zemmer kevin.zemmer@ifor.math.ethz.ch
Christoph Glanzer christoph.glanzer@ifor.math.ethz.ch

IFOR
Institute for Operations Research
Department of Mathematics

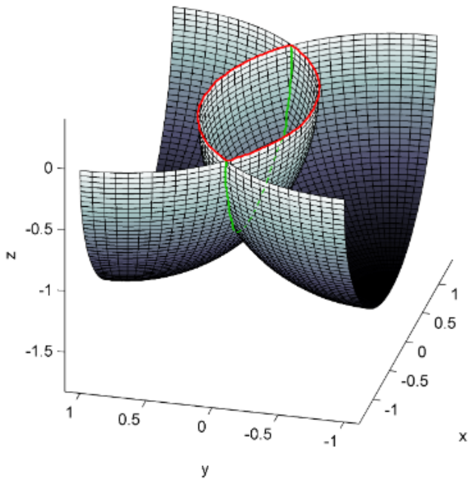
Mathematical Optimization — Assignment 8

https://moodle-app2.let.ethz.ch/course/view.php?id=3610

Exercise 1: Subgradients

$$\begin{aligned} f_1(x,y) = &\quad -\sqrt{2-x^2-(y-1)^2}, &\quad \mathrm{dom} f_1 = \{(x,y) \mid x^2+(y-1)^2 \le 2\}, \\ f_2(x,y) = &\quad -\sqrt{2-x^2-(y+1)^2}, &\quad \mathrm{dom} f_2 = \{(x,y) \mid x^2+(y+1)^2 \le 2\}, \\ f(x,y) = &\quad \max\{f_1(x,y), f_2(x,y)\}, &\quad \mathrm{dom} f = \mathrm{dom} f_1 \cap \mathrm{dom} f_2 \end{aligned}$$

Let us first visualize the situation with a picture ($\mathrm{dom} f$ is the area delimited by the red curve, projected onto the xy-plane):



We recall a rule for computing subgradients of the maximum of convex functions seen in the lecture:

$$f(x) := \max_i f_i(x) \Longrightarrow \partial f(x_0) = \mathrm{conv}\{\partial f_i(x_0) \mid f_i(x_0) = f(x_0)\}, \forall x_0 \in \bigcap_i \mathrm{dom}(f_i).$$

- The subgradient of f at the boundary points $\{(x,y) \in \mathrm{dom} f \mid f(x,y) = 0\}$ (red line in the picture above) is empty, since any supporting hyperplane of f at those points should be vertical, while the subgradient is constructed using the non-vertical supporting hyperplanes of epi(f).
- The subgradient of f at the points $\{(x,y) \in \mathrm{dom} f \mid y \neq 0, f(x,y) \neq 0\}$ is a singleton, since the function there is differentiable (it consists of f_1 or f_2). The subdifferential of f is then equal to the differential of f_1 resp. f_2 (obtained by computing $\partial f_i(x,y)/\partial x$ and $\partial f_i(x,y)/\partial y$):

  If $y < 0$ (i.e. $f = f_1$): $\nabla f(x,y) = \frac{1}{\sqrt{2-x^2-(y-1)^2}} \binom{x}{y-1}$.

  If $y > 0$ (i.e. $f = f_2$): $\nabla f(x,y) = \frac{1}{\sqrt{2-x^2-(y+1)^2}} \binom{x}{y+1}$.
- The subgradient of f at the points $\{(x,0) \in \mathrm{dom} f \mid f(x,y) \neq 0\} = \{(x,0) \mid -1 < x < 1\}$ (i.e. where f is given by both f_1 and f_2, the green line in the picture above) is, according to the rule above,

the convex hull of the (sub)gradients of f_1 and f_2:

$$\begin{aligned}\partial f(x,0) = &\ \operatorname{conv}\left\{\frac{1}{\sqrt{1-x^2}}\begin{pmatrix}x\\-1\end{pmatrix}, \frac{1}{\sqrt{1-x^2}}\begin{pmatrix}x\\+1\end{pmatrix}\right\}\\ =&\ \left\{\lambda\frac{1}{\sqrt{1-x^2}}\begin{pmatrix}x\\-1\end{pmatrix} + (1-\lambda)\frac{1}{\sqrt{1-x^2}}\begin{pmatrix}x\\+1\end{pmatrix}\,\middle|\, \lambda\in[0,1]\right\}\\ =&\ \left\{\frac{1}{\sqrt{1-x^2}}\begin{pmatrix}x\\t\end{pmatrix}\,\middle|\, t\in[-1,+1]\right\}.\end{aligned}$$

Exercise 2: Separation of Convex Sets

(a) We consider only the case of strong separation, the case of separation is similar, but simpler.

*Proof.*To prove the convexity of $C-D$, let $x-y$, $z-w \in C-D$, where $x,z\in C$, $y,w\in D$ and $0\le\lambda\le 1$. Then

$(1-\lambda)(x-y)+\lambda(z-w) = (1-\lambda)x+\lambda z-(1-\lambda)y+\lambda w \in C-D$

by convexity of C and D. Thus $C-D$ is convex.

$(i) \Rightarrow (ii)$

Let $S=\{x:\alpha\le u\cdot x\le\beta\}$, $\alpha<\beta$, be a set which strongly separates C and D, say $C\subseteq\{x: u\cdot x\le\alpha\}$ and $D\subseteq\{y: u\cdot y\ge\beta\}$. Then $C-D\subseteq\{x-y: u\cdot(x-y)\le\alpha-\beta\}$. Thus $C-D$ and $\{0\}$ are separated by the set $\{z:\alpha-\beta\le u\cdot z\le 0\}$.

$(ii) \Rightarrow (i)$

Let the set $S=\{z:-\gamma\le u\cdot z\le 0\}$, $\gamma>0$, separate $\{0\}$ and $C-D$. Then $u\cdot(x-y)\le-\gamma$, i.e. $u\cdot x+\gamma\le u\cdot y$ for all $x\in C$ and $y\in D$. Let $\alpha = sup\{u\cdot x: x\in C\}$. Then the set $\{z:\alpha\le u\cdot z\le\alpha+\gamma\}$ separates C and D.

□

(b) *Proof.* Choose $p\in C, q\in D$ having minimum distance. Let $u=q-p(\neq 0)$. Then the set $\{x: u\cdot p\le u\cdot x\le u\cdot q\}$ separates C and D.

□

Exercise 3: Equality constrained least squares

Recall the KKT-conditions from the lecture for $\min\{f(x)\mid g(x)\le b\}$ and $x^*\in\operatorname{dom}(f)$:

There exist $\lambda_i^*\ge 0$ for all $i=1,\dots,m$ such that

- $\lambda_i^*=0$ for all $i\notin I(x^*)=\{j\colon g_i(x^*)=b_i\}$
- $\nabla f(x^*)+\sum_{i=1}^m\lambda_i^*\nabla g_i(x^*)=0.$

In this example, these conditions are

$$2A^T(Ax^*-b)+G^T\nu^*=0 \qquad Gx^*=h.$$

Assignment 09

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institute for Operations Research
ETH Zurich HG G21-22

Kevin Zemmer kevin.zemmer@ifor.math.ethz.ch
Christoph Glanzer christoph.glanzer@ifor.math.ethz.ch

IFOR
Institute for Operations Research
Department of Mathematics

Mathematical Optimization — Solution 9

https://moodle-app2.let.ethz.ch/course/view.php?id=2180

Exercise 1: Unimodular and Totally Unimodular Matrices
Let $A \in \mathbb{R}^{m\times n}$ and let I be the n-dimensional identity matrix.

a) $\Rightarrow$: Assume that A is totally unimodular. Choose any square submatrix $(A')^{\mathrm{T}}$ of A^{T}. Then, A' is a square submatrix of A and therefore, $\det{(A')}^{\mathrm{T}} = \det{(A')} \in \{\pm 1, 0\}$.
$\Leftarrow$: Apply the statement that we proved above to A^{T}.

b) Assume that A is totally unimodular. Choose any square submatrix $B \in \mathbb{R}^{k\times k}$ of $[A \mid I]$. B arises from $[A \mid I]$ by removing some rows and columns. Therefore, $B = [A' \mid I']$, where $A' \in \mathbb{R}^{k\times l}$ and $I' \in \mathbb{R}^{k\times (k-l)}$. Applying Laplace's rule to all columns in I' consecutively erases the corresponding rows in A' (rows corresponding to the unit vector forming the respective column in I') and leads to $\det(B) = \pm\det(B')$, where B' is a sub-matrix of A' and therefore also of A. Thus, $\det(B) \in \{\pm 1, 0\}$.

c) Assume that $[A \mid I]$ is unimodular. Choose any quadratic submatrix $A' \in \mathbb{R}^{k\times k}$ of A. A' arises from A by removing certain columns and rows. We will now choose a $n \times n$ submatrix B from $[A \mid I]$ such that $\det(B) = \pm\det(A')$. First, let B contain all the columns of A which are left in A', which are k in total. Concerning the rows, we removed $m - k$ rows from A to get to A'. Choose the rest of the columns of B to be the columns of I corresponding to the indices of the removed rows. We therefore choose $m - k + k$ columns in total, making B a $m \times m$ matrix and a submatrix of $[A \mid I]$. Therefore, $\det(B) \in \{\pm 1, 0\}$. When calculating the determinant of B, we see when using Laplace on the $m-k$ index vectors (column-wise), that $\det(B) = \det(A')$ as we eliminate the exact same rows. Therefore, $\det(A') \in \{0, \pm 1\}$, proving the statement.

d) " $\Leftarrow$ ": Choose a submatrix A' of A. Then, A' is also a submatrix of $\begin{bmatrix} A \\ -A \end{bmatrix}$, leading to $\det(A') \in \{\pm 1, 0\}$.
" $\Rightarrow$ ": Choose any submatrix B of $\begin{bmatrix} A \\ -A \end{bmatrix}$. Either there are indices i, j such that $B_{i,\cdot} = -B_{j,\cdot}$ in which case the determinant of B is zero or otherwise, B arises from a submatrix of A by multiplying some of its rows by (-1). Therefore, $\det(B) \in \{0, \pm 1\}$.

e) From b), c) and d), we know that
$$A \text{ TU} \Leftrightarrow A^{\mathrm{T}} \text{ TU} \Leftrightarrow [A^{\mathrm{T}} \mid I] \text{ TU} \Leftrightarrow [A^{\mathrm{T}} \mid I \mid -A^{\mathrm{T}} \mid -I] \text{ TU}.$$
Permuting the columns does not change any determinant of any submatrix, therefore
$$[A^{\mathrm{T}} \mid I \mid -A^{\mathrm{T}}- \mid I] \text{ TU} \Leftrightarrow [A^{\mathrm{T}} \mid -A^{\mathrm{T}} \mid I \mid -I] \text{ TU}.$$
Now, as A TU $\Leftrightarrow A^{\mathrm{T}}$ TU, $[A^{\mathrm{T}} \mid -A^{\mathrm{T}} \mid I \mid -I]$ TU $\Leftrightarrow [A^{\mathrm{T}} \mid -A^{\mathrm{T}} \mid I \mid -I]^{\mathrm{T}} = \begin{bmatrix} A \\ -A \\ I \\ -I \end{bmatrix}$ is totally unimodular.

f) We use the Theorem by Ghouila-Houri on the columns of A. Let $R \subseteq \{1, \dots, n\}$ be a subset of the columns. Now, the matrix $A_{\cdot,R}$ is again consecutive-ones, which is why we can w.l.o.g. assume that $R = \{1, \dots, n\}$. Now, we create our partition $R = R_1 \cup R_2$ such that R_1 contains all even numbers of indices and R_2 contains all odd numbers. Consider now an arbitrary row $i \in \{1, \dots, m\}$. It holds that $\sum_{j\in R_1} A_{i,j} - \sum_{j\in R_2} A_{i,j} \in \{0, \pm 1\}$ as the matrix is a consecutive-ones matrix and we sum up the elements alternately. This proves the statement by Ghouila-Houri's criterion.

g) The matrix stays TU. This can be seen by Ghouila-Houri's criterion as follows:

Let $R \subseteq \{1, \dots, m+1\}$ be a subset of rows. We do a case by case analysis, depending on whether the $(n+1)^{\text{st}}$ row is part of R or not.

Assume first that $R \subseteq \{1, \dots, m\}$. We assume that G is bipartite, i.e. we can partition $V = V_1 \cup V_2$ such that $V_1 \cap V_2 = \emptyset$ and all edges run between V_1 and V_2. In this case, partition the rows according to whether the vertex corresponding to the row is in V_1 or V_2. Call the partition $R = R_1 \cup R_2$. We know therefore that

$$\sum_{i \in R_1} A_{i,\cdot} - \sum_{i \in R_2} A_{i,\cdot} \in \{0, 1, -1\}^n,$$

in every column, we cannot have more than one entry in R_i, $i = 1, 2$.

In the other case, $\{m+1\} \in R$. Choose now $R_1 := R \setminus \{m+1\}$, $R_2 := \{m+1\}$. Consider any column $j \in \{1, \dots, n\}$. Then,

$$\sum_{i \in R_1} A_{i,j} - \sum_{i \in R_2} A_{i,j} = \underbrace{\sum_{i \in R \setminus \{m+1\}} A_{i,j}}_{\in \{0,1,2\}} - \underbrace{A_{m+1,j}}_{1} \in \{0, 1, -1\}.$$

As this holds for every column, the claim follows by Ghouila-Houri's criterion.

h) Let $J \subseteq \{1, \dots, n_1 + n_2\}$. Define $J_B = J \cap \{n_1 + 1, \dots, n_1 + n_2\}$, $J_A = J \cap \{1, \dots, n_1\}$. Let J_B^+, J_B^- be a partitions for $\begin{bmatrix} b^{\mathrm{T}} \\ B \end{bmatrix}$ according to the Ghouila-Houri Theorem.

- If $\sum_{i \in J_B^+} b_i - \sum_{i \in J_B^-} b_i = 0$, then we can choose a partition of J_A for $[A]$ into J_A^+, J_A^- according to the G.H.-Theorem. Then $J_A^+ \cup J_B^+$ and $J_A^- \cup J_B^-$ give the desired decomposition of J into two sets.
- Otherwise, w.l.o.g. $\sum_{i \in J_B^+} b_i - \sum_{i \in J_B^-} b_i = 1$. Then choose a partition of $J_A \cup \{n_1 + 1\}$ for $[A \mid a]$ into $J_A^+ \cup \{n_1 + 1\}$, J_A^- according to the G.H.-Theorem. Then $J_A^+ \cup J_B^+$ and $J_A^- \cup J_B^-$ give the desired decomposition of J.

Exercise 2: A convex problem in which strong duality fails

(a) $f(x_1, x_2) := e^{-x_1}$ and $g(x_1, x_2) := \frac{x_1^2}{x_2}$ are convex on the domain $D := \{(x_1, x_2) \mid x_2 > 0\}$ since they are smooth and their Hessian is positive semidefinite,

$$\nabla^2 f(x_1, x_2) = \begin{pmatrix} e^{-x_1} & 0 \\ 0 & 0 \end{pmatrix} \succeq 0 \qquad \nabla^2 g(x_1, x_2) = \begin{pmatrix} \frac{2}{x_2} & -2\frac{x_1}{x_2^2} \\ -2\frac{x_1}{x_2^2} & 2\frac{x_1^2}{x_2^3} \end{pmatrix} \succeq 0$$

Hence, the problem is a convex optimization problem. The feasible set is $\{(0, x_2) \in \mathbb{R}^2 \mid x_2 > 0\}$ and the optimal value is $p^* = 1$.

(b) The Lagrangian is $\inf_{x \in D} L(x, u)$, where $L(x, u) = e^{-x_1} + u x_1^2 / x_2$. Then we have

$$\inf_{x \in D} L(x, u) = \inf_{x \in D} \{e^{-x_1} + u x_1^2 / x_2\} = \begin{cases} 0 & u \geq 0 \\ -\infty & u < 0 \end{cases}$$

(note that for $u < 0$ the function $L(x, u)$ is decreasing in x). Thus, we can write the dual problem as

$$\begin{array}{ll} \text{maximize} & 0 \\ \text{subject to} & u \geq 0 \end{array}$$

with optimal value $d^* = 0$. The optimal duality gap is $p^* - d^* = 1$.

(c) Slater's condition is not satisfied. Namely, no point $(x_1, x_2) \in D$ can satisfy $x_1^2 / x_2 < 0$.

Assignment 10

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institute for Operations Research
ETH Zurich HG G21-22

Stephan Artmann stephan.artmann@ifor.math.ethz.ch
Christoph Glanzer christoph.glanzer@ifor.math.ethz.ch

IFOR
Institute for Operations Research
Department of Mathematics

Mathematical Optimization — Solution 10

https://moodle-app2.let.ethz.ch/course/view.php?id=3610

Exercise 1: The pigeonhole principle as LP-formulation

i) We want to show that $(P_2^R) \subseteq (P_1^R)$. We therefore need to show that all constraints of the latter are implied by constraints of the former. All constraints apart from $x_{ij} + x_{kj} \leq 1$ for $j = 1, \ldots, n$ and $i, k = 1, \ldots, n+1,\ i \neq k$ are fulfilled. This set of constraints is fulfilled as well, as $x_{ij} + x_{kj} \leq \sum_{i=1}^{n+1} x_{ij} \leq 1$, using that $x_{ij} \geq 0$.

ii) Model $(P1)$ does not have an empty LP-relaxation, as can be seen by letting $x_{ij} = \frac{1}{n}$, for all $i = 1, \ldots, n+1$ and $j = 1, \ldots, n$: It holds that $\sum_{j=1}^{n} x_{ij} = n/n = 1$ for all $i = 1, \ldots, n+1$ and $x_{ij} + x_{jk} = \frac{2}{n} \leq 1$ as we assumed that $n \geq 2$.

We want to show that the LP-relaxation of the polyhedron in model (P_2) is infeasible. The LP-relaxation is given by:

$$(P_2^R) \qquad \begin{array}{ll} \sum_{j=1}^{n} x_{ij} = 1, & i = 1, \ldots, n+1, \\ \sum_{i=1}^{n+1} x_{ij} \leq 1, & j = 1, \ldots, n, \\ x_{ij} \geq 0, & i = 1, \ldots, n+1,\ j = 1, \ldots, n. \end{array}$$

Note that we removed the constraints of the form $x_{ij} \leq 1$ as these are implied by the second constraint already. We want to use Farkas' Lemma. This amounts to calculating the dual of (P_2^R) with $c = \vec{0}$. Define y_i, $i = 1, \ldots, n+1$, corresponding to the first type of constraint of (P_2^R), z_j, $j = 1, \ldots, n$ for the second type of constraint of (P_2^R) and w_{ij} for the non-negativity constraints. We obtain the following.

$$\min \left\{ \sum_{i=1}^{n+1} y_i + \sum_{j=1}^{n} z_j \;\middle|\; y_i + z_j + w_{ij} \geq 0,\ \forall i, j,\ y_i \text{ free},\ z_j \geq 0,\ w_{ij} \leq 0 \right\}.$$

Choose $y_i = -1$, $z_j = 1$, $w_{ij} = 0$ for all i and j. This forms a feasible solution with objective value $-(n+1) + n = -1$, implying that P_2^R is infeasible.

Exercise 2: Knapsack: Bad approximation of optimal solution

i) Consider the following instance of a binary knapsack problem, where $N \geq 1$, $N \in \mathbb{N}$:

$$\begin{array}{ll} \max & x_1 + N x_2 \\ \text{s.t.} & \frac{1}{N+1} x_1 + x_2 \leq 1, \\ & x \in \{0,1\}^2. \end{array}$$

It holds that $\frac{c_1}{a_1} \geq \frac{c_2}{a_2}$. In this case, $x^1 = (1,0)^{\mathrm{T}}$, while $x^2 = (0,1)^{\mathrm{T}}$. By checking all possible solutions, we can easily see that $x^2 = x^*_{BK}$, where $c^{\mathrm{T}} x^2 = N$, but $c^{\mathrm{T}} x^1 = 1$. For $N \to \infty$, we can get an arbitrarily bad approximation of $c^{\mathrm{T}} x^*_{BK}$.

ii) Consider the following instance of a binary knapsack problem, where $N \geq 1$, $N \in \mathbb{N}$:

$$\begin{array}{ll} \max & N x_1 + x_2 \\ \text{s.t.} & x_1 + 2x_2 \leq 2, \\ & x \in \{0,1\}^2. \end{array}$$

It holds that $\frac{c_1}{a_1} \geq \frac{c_2}{a_2}$. Again, $x^1 = (1,0)^{\mathrm{T}}$, while $x^2 = (0,1)^{\mathrm{T}}$. By checking all possible solutions, we can easily see that $x^1 = x^*_{BK}$, where $c^{\mathrm{T}}x^1 = N$, but $c^{\mathrm{T}}x^2 = 1$. For $N \to \infty$, we can get an arbitrarily bad approximation of $c^{\mathrm{T}}x^*_{BK}$.

Exercise 3: Integral Polyhedra
Let us start by proving the hint:

Claim:
Let P be a (rational) polyhedron and v a vertex of P. Then, there exists $c \in \mathbb{Q}^n$ such that $c^{\mathrm{T}}v > c^{\mathrm{T}}w$ for $w \neq v$ being a vertex of P.

Proof: Choose $c^{\mathrm{T}} := \sum_{i \in I(v)} A_{i,\cdot} \in \mathbb{Q}^n$ (as P is a rational polyhedron). Let $w \in P$ be a different vertex of P. Then, there exists $i \in I(v)$ such that $A_{i,\cdot}w < b_i$. Therefore, $c^{\mathrm{T}}w < \sum_{i \in I(v)} A_{i,\cdot}v = \sum_{i \in I(v)} b_i$. $\square$

" $\Rightarrow$ ": Suppose that P is a non-empty, pointed, integral polyhedron, i.e. every vertex of P is integral. Assume that for a given $c \in \mathbb{Z}^n$, the problem is bounded. Then, there exists an optimal solution x^* in P which is a vertex and therefore integral.

" $\Leftarrow$ ": Assume by contradiction that P is not integral, i.e. there exists a fractional vertex x^* of P. Let x^*_j be a fractional component of x^*. Since x^* is a vertex of P, we know that there exists $c \in \mathbb{Q}^n$ such that x^* is the unique optimal vertex solution of $\max\{c^{\mathrm{T}}x \mid x \in P\}$.

Now, scale c such that it is in $\mathbb{Z}^n$. Then, for $a \in \mathbb{Z}$ sufficiently large and $\bar{c} = c + (1/a)e_j$, x^* also maximizes $\bar{c}$. Thus, x^* is an optimal solution for the integer cost vectors ac and $a\bar{c}$. Since $a\bar{c}^T x^* - ac^T x^* = x^*_j$, either $a\bar{c}^T x^*$ or $ac^T x^*$ is fractional, implying that there exists an integer vector for which the optimal solution is fractional, a contradiction.

Assignment 11

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institute for Operations Research
ETH Zurich HG G21-22

Stephan Artmann stephan.artmann@ifor.math.ethz.ch
Christoph Glanzer christoph.glanzer@ifor.math.ethz.ch

IFOR
Institute for Operations Research
Department of Mathematics

Mathematical Optimization — Solution 11

https://moodle-app2.let.ethz.ch/course/view.php?id=3610

Exercise 1: Mixed-integer feasibility with a fixed number of integer variables
Define $P_{MI} := \{\binom{x}{y} \in \mathbb{Z}^d \times \mathbb{R}^n \mid Ax + By \leq c\}$ and let P be its LP-relaxation. Our goal is to decide whether $P_{MI} \neq \emptyset \Leftrightarrow P \cap (\mathbb{Z}^d \times \mathbb{R}^n) \neq \emptyset$.

Consider $\text{proj}_x P = \{x \in \mathbb{R}^d \mid \exists y \in \mathbb{R}^n \text{ s.t. } Ax + By \leq c\}$. From the lecture, we know that $P = \emptyset \Leftrightarrow \text{proj}_x P = \emptyset$. Additionally, it holds that $P_{MI} = \emptyset \Leftrightarrow (\text{proj}_x P) \cap \mathbb{Z}^d = \emptyset$:

"$\Rightarrow$:" Follows directly as an element in $(\text{proj}_x P) \cap \mathbb{Z}^d$ would be a contradiction.

"$\Leftarrow$:" This implies that $\{x \in \mathbb{Z}^d \mid \exists y \in \mathbb{R}^n \text{ s.t. } Ax + By \leq c\} = \emptyset$, which implies that $P_{MI} = \emptyset$.

By applying the IP-feasibility oracle, we can find an $\bar{x} \in (\text{proj}_x P) \cap \mathbb{Z}^d$ if it exists or decide that no such x exists, which is equivalent to $P_{MI} = \emptyset$. If it exists, the existence of $\bar{x}$ implies that there is a corresponding y such that $\binom{\bar{x}}{y} \in P_{MI}$. In order to find such a y, fix the entries of $\bar{x}$ and apply the LP-feasibility oracle to the following LP: $\{y \in \mathbb{R}^n \mid A\bar{x} + By \leq c\} = \{y \in \mathbb{R}^n \mid By \leq c - A\bar{x}\}$.

Exercise 2: Matching and Perfect Matching Polytope
We claim that $PM'(G) = PM(G)$.

Note that in the case where $|V|$ is odd, $PM'(G) = \emptyset = PM(G)$, as $x(E[V]) = x(E) = \frac{|V|}{2} > \frac{|V|-1}{2}$ for $PM'(G)$ and $x(\delta(V)) = x(\emptyset) = 0 < 1$ for $PM(G)$. Thus, we can assume w.l.o.g. that $|V|$ is even.

"$\subseteq$": Let $x \in$ LHS.

- We know that $2x(E) = \sum_{v \in V} \underbrace{x(\delta(v))}_{\leq 1} \leq |V| = 2x(E) \Rightarrow x(\delta(v)) = 1$.
- Let $S \subseteq V$, $|S|$ odd, $S \neq \emptyset$. Then, $|S^C|$ also has to be odd as $|V| = |S| + |S^C|$ is even. Thus,

$$x(E[S]) + x(E[S^C]) \leq \frac{|S|-1}{2} + \frac{|S^C|-1}{2} = \frac{|V|-2}{2} = \frac{|V|}{2} - 1.$$

  As furthermore $\delta(S) = E \setminus \big(E[S] \cup E[S^C]\big)$, it holds that

$$x(\delta(S)) = x(E) - \big(x(E[S]) + x(E[S^C])\big) \geq \frac{|V|}{2} - \frac{|V|}{2} + 1 = 1.$$

"$\supseteq$": Let $x \in$ RHS.

- $x(\delta(v)) \leq 1$ is clearly satisfied for all $v \in V$.
- It holds that $2x(E) = \sum_{v \in V} x(\delta(v)) = |V| \Rightarrow x(E) = \frac{|V|}{2}$.
- Let $S \subseteq V$, $|S|$ odd. Then,

$$|S| = \sum_{v \in S} |\delta(v)| = 2 \sum_{e \in E[S]} x(e) + \sum_{e \in \delta(S)} x(e).$$

  Therefore,

$$x(E[S]) = \frac{|S| - x(\delta(S))}{2} \leq \frac{|S|-1}{2},$$

  as $x(\delta(S)) \geq 1$.

Exercise 3: LP Solution and the Normal Cone

We assume that $x^* := \max\{c^{\mathrm{T}}x \mid x \in P\}$ exists and is finite, from which follows that also $y^* := \min\{y^{\mathrm{T}}b \mid y^{\mathrm{T}}A = c^{\mathrm{T}},\ y \geq 0\}$ exists and is finite. Using complementary slackness, we know that x^* is optimal for the primal problem and y^* is optimal for the dual problem if and only if $y_i^*\,(A_{i,\cdot}x^* - b_i) = 0$, for all $i \in \{1, \ldots, m\}$. Let I be the constraints where x^* is tight. The previous argument proves that $c \in \operatorname{cone}(\{A_{i,\cdot},\ i \in I\})$ as it gives us a valid combination: We know that $y_i^* = 0$ for $i \notin I$ and therefore, $(y^*)^{\mathrm{T}}A = \sum_{i\in I} y_i^* A_{i,\cdot} = c^{\mathrm{T}}$ and $y^* \geq 0$, from which it follows that $c \in \operatorname{cone}\left(\left\{A_{i,\cdot}^{\mathrm{T}},\ i \in I\right\}\right)$.

Assignment 12

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institute for Operations Research
ETH Zurich HG G21-22

Stephan Artmann stephan.artmann@ifor.math.ethz.ch
Christoph Glanzer christoph.glanzer@ifor.math.ethz.ch

IFOR
Institute for Operations Research
Department of Mathematics

Mathematical Optimization — Solution 12

https://moodle-app2.let.ethz.ch/course/view.php?id=2180

Exercise 1: Submodular Functions

(a) "$\Rightarrow$": Let $S := A \cup \{j\}$, $T := A \cup \{k\}$. Then,

$$f(S \cup T) - f(T) \le f(S) - f(S \cap T) \quad \Leftrightarrow \quad f(A \cup \{j,k\}) - f(A \cup \{k\}) \le f(A \cup \{j\}) - f(A).$$

"$\Leftarrow$": Let $A, B \subseteq N$, $S := A \cap B$, $A \setminus B := \{j_1, \ldots, j_r\}$, $B \setminus A := \{k_1, \ldots, k_s\}$. Then,

$$\begin{aligned}
f(B) - f(A \cap B) &= f(S \cup \{k_1, \ldots, k_s\}) - f(S) \\
&= \sum_{i=1}^{s} f(S \cup \{k_1, \ldots, k_i\}) - f(S \cup \{k_1, \ldots, k_{i-1}\}) \\
&\ge \sum_{i=1}^{s} f(S \cup \{k_1, \ldots, k_i\} \cup \{j_1\}) - f(S \cup \{k_1, \ldots, k_{i-1}\} \cup \{j_1\}) \\
&\ge \ \ldots \ \ge \sum_{i=1}^{s} f(S \cup \{k_1, \ldots, k_i\} \cup \{j_1, \ldots, j_r\}) - f(S \cup \{k_1, \ldots, k_{i-1}\} \cup \{j_1, \ldots, j_r\}) \\
&= \sum_{i=1}^{s} f(A \cup \{k_1, \ldots, k_i\}) - f(A \cup \{k_1, \ldots, k_{i-1}\}) = f(A \cup B) - f(A).
\end{aligned}$$

$\square$

(b) "$\Rightarrow$": Let $S, T \subseteq N$, $T \setminus S := \{j_1, \ldots, j_r\}$. Then,

$$\begin{aligned}
f(T) &\overset{\text{non-decreasing}}{\le} f(S \cup T) = f(S) + (f(S \cup T) - f(S)) \\
&= f(S) + \sum_{i=1}^{r} (f(S \cup \{j_1, \ldots, j_i\}) - f(S \cup \{j_1, \ldots, j_{i-1}\})) \\
&\overset{\text{(a)}}{\le} f(S) + \sum_{i=1}^{r} (f(S \cup \{j_i\}) - f(S)).
\end{aligned}$$

"$\Leftarrow$": Setting $T := S \cup \{j, k\}$ gives

$$\begin{aligned}
& f(S \cup \{j,k\}) \le f(S) + f(S \cup \{j\}) - f(S) + f(S \cup \{k\}) - f(S) \\
\Leftrightarrow\ & f(S \cup \{j\}) - f(S) \ge f(S \cup \{j,k\}) - f(S \cup \{k\}),
\end{aligned}$$

which is (a) and therefore implies submodularity.

For monotonicity, let $T \subseteq S$. Now,

$$f(T) \le f(S) + \sum_{j \in T \setminus S} f(S \cup \{j\}) - f(S) = f(S).$$

$\square$

Exercise 2: Matroid-union

Let $I, J \in \mathcal{F}$ with $|I| < |J|$. Without loss of generality, assume that $|I \cap E_1| < |J \cap E_1|$ (otherwise, exchange the roles of M_1 and M_2). Using the hint, there exists $e \in (J \cap E_1) \setminus (I \cap E_1)$ such that $(I \cap E_1) \cup e \in \mathcal{F}_1$. Therefore, $I \cup \{e\} \in \mathcal{F}$, as $(I \cap E_1) \cup e = I \cup e \cap E_1 \in \mathcal{F}_1$ and $(I \cup e) \cap E_2 = I \cap E_2 \in \mathcal{F}_2$, where we used that $e \in E_1 \Rightarrow e \notin E_2$.

Exercise 3: Hamiltonian Paths and Matroid Intersection

We consider the following three matroids for intersection:

$\mathcal{M}_1$: **Graphic matroid** $\mathcal{M}_1 = (\mathcal{F}_1, E)$. **Note:** We ignore the direction of the edges.

$\mathcal{M}_2$: **Partition matroid** $\mathcal{M}_2 = (\mathcal{F}_2, E)$ where each node $v \in V$ has at most one outgoing edge: $\mathcal{F}_2 = \{A \subseteq E \quad | \forall v \in V : |\{(v, w) \mid (v, w) \in A\}| \leq 1\}$

$\mathcal{M}_3$: **Partition matroid** $\mathcal{M}_3 = (\mathcal{F}_3, E)$ where each node $v \in V$ has at most one incoming edge: $\mathcal{F}_3 = \{A \subseteq E \quad | \forall v \in V : |\{(w, v) \mid (w, v) \in A\}| \leq 1\}$

Any independent set in $\mathcal{F}_1 \cap \mathcal{F}_2 \cap \mathcal{F}_3$ represents a disjoint union of directed, simple paths. Therefore, G admits a directed Hamiltonian path if and only if a maximally independent set of cardinality $|V| - 1$ exists in $\mathcal{F}_1 \cap \mathcal{F}_2 \cap \mathcal{F}_3$. Note that in the case where there doesn't exist a directed path between any two vertices of G, an independent set of cardinality $|V| - 1$ cannot exist. Determining whether a Hamiltonian path (or hamiltonian cycle) exists in a given graph is $\mathcal{NP}$-complete. Hence finding the maximal independent set in the intersection of three matroids is $\mathcal{NP}$-hard.

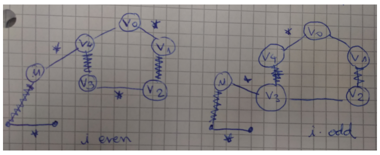


Figure 9: If i is even: replace supernode C in P by the sequence $v_i, v_{i-1}, \ldots, v_0$. This gives M-augmenting path in G. If i is odd: replace supernode C in P' by the sequence $v_i, v_{i+1}, \ldots, v_{t-1}, v_0$. This gives M-augmenting path in G.

Glossary

Term	**Brief explanation of the term**
$A_{i\cdot}$	row access on a matrix A
$A_{\cdot j}$	column access on a matrix A
bfs	basic feasible solution
BFS	Breadth-first search
DP	dynamic programming
ERO	elementary row operation
KMC	Klee Mint Cube
PSD	Positive semi definite
SA	Simplex Algorithm
st	such that
$wlog$	without loss of generality

TODO

This is the chapter on what still has to be improved in the summary. Please update this list by writing the todos directly into the summary where necessary by writing:

```
\todo{Something that still has to be done}
```

For a missing figure, please write:

```
\missingfigure{A missing figure}
```

Everyone that picks up the work on this summary will thank you for keeping this up-to-date properly.

Todo list