Part II

Linear Programming

Brewery brews ale and beer.

- Production limited by supply of corn, hops and barley malt
- Recipes for ale and beer require different amounts of resources

Brewery brews ale and beer.

- Production limited by supply of corn, hops and barley malt
- Recipes for ale and beer require different amounts of resources

Brewery brews ale and beer.

- Production limited by supply of corn, hops and barley malt
- Recipes for ale and beer require different amounts of resources

	Corn (kg)	Hops (kg)	Malt (kg)	Profit (€)
ale (barrel)	5	4	35	13
beer (barrel)	15	4	20	23
supply	480	160	1190	

	Corn (kg)	Hops (kg)	Malt (kg)	Profit (€)
ale (barrel)	5	4	35	13
beer (barrel)	15	4	20	23
supply	480	160	1190	

How can brewer maximize profits?

	Corn (kg)	Hops (kg)	Malt (kg)	Profit (€)
ale (barrel)	5	4	35	13
beer (barrel)	15	4	20	23
supply	480	160	1190	

How can brewer maximize profits?

- only brew ale: 34 barrels of ale
- only brew beer: 32 barrels of beer
- > 7.5 barrels ale, 29.5 barrels beer
- ▶ 12 barrels ale, 28 barrels beer

	Corn (kg)	Hops (kg)	Malt (kg)	Profit (€)
ale (barrel)	5	4	35	13
beer (barrel)	15	4	20	23
supply	480	160	1190	

How can brewer maximize profits?

only brew ale: 34 barrels of ale

⇒ 442€

- only brew beer: 32 barrels of beer
- ▶ 7.5 barrels ale, 29.5 barrels beer
- 10 homele de 20 homele hoor
- ▶ 12 barrels ale, 28 barrels beer

	Corn (kg)	Hops (kg)	Malt (kg)	Profit (€)
ale (barrel)	5	4	35	13
beer (barrel)	15	4	20	23
supply	480	160	1190	

How can brewer maximize profits?

only brew ale: 34 barrels of ale

⇒ 442€

only brew beer: 32 barrels of beer

→ /36 €

▶ 7.5 barrels ale, 29.5 barrels beer

 \Rightarrow 776 \in

▶ 12 barrels ale, 28 barrels beer

⇒ 800€

	Corn (kg)	Hops (kg)	Malt (kg)	Profit (€)
ale (barrel)	5	4	35	13
beer (barrel)	15	4	20	23
supply	480	160	1190	

How can brewer maximize profits?

only brew ale: 34 barrels of ale

only brew beer: 32 barrels of beer ⇒ 736 €

> 7.5 barrels ale, 29.5 barrels beer

► 12 barrels ale, 28 barrels beer ⇒ 800 €

⇒ 442 €

	Corn (kg)	Hops (kg)	Malt (kg)	Profit (€)
ale (barrel)	5	4	35	13
beer (barrel)	15	4	20	23
supply	480	160	1190	

How can brewer maximize profits?

only brew ale: 34 barrels of ale

⇒ 442€

only brew beer: 32 barrels of beer

⇒ 736€

▶ 7.5 barrels ale, 29.5 barrels beer

⇒ 776 €

▶ 12 barrels ale, 28 barrels beer

	Corn (kg)	Hops (kg)	Malt (kg)	Profit (€)
ale (barrel)	5	4	35	13
beer (barrel)	15	4	20	23
supply	480	160	1190	

How can brewer maximize profits?

only brew ale: 34 barrels of ale ⇒ 442 €

only brew beer: 32 barrels of beer ⇒ 736 €

7.5 barrels ale, 29.5 barrels beer ⇒ 776€

▶ 12 barrels ale, 28 barrels beer

	Corn (kg)	Hops (kg)	Malt (kg)	Profit (€)
ale (barrel)	5	4	35	13
beer (barrel)	15	4	20	23
supply	480	160	1190	

How can brewer maximize profits?

only brew ale: 34 barrels of ale

⇒ 442€

only brew beer: 32 barrels of beer

⇒ 736€

▶ 7.5 barrels ale, 29.5 barrels beer

⇒ 776 €

▶ 12 barrels ale, 28 barrels beer

⇒ 800€

	Corn (kg)	Hops (kg)	Malt (kg)	Profit (€)
ale (barrel)	5	4	35	13
beer (barrel)	15	4	20	23
supply	480	160	1190	

How can brewer maximize profits?

only brew ale: 34 barrels of ale ⇒ 442 €

only brew beer: 32 barrels of beer ⇒ 736€

7.5 barrels ale, 29.5 barrels beer ⇒ 776€

▶ 12 barrels ale, 28 barrels beer ⇒ 800€

max
$$13a + 23b$$

s.t. $5a + 15b \le 480$
 $4a + 4b \le 160$
 $35a + 20b \le 1190$
 $a, b \ge 0$

- Introduce variables *a* and *b* that define how much ale and beer to produce.
- Choose the variables in such a way that the objective function (profit) is maximized.
- Make sure that no constraints (due to limited supply) are violated.

max
$$13a + 23b$$

s.t. $5a + 15b \le 480$
 $4a + 4b \le 160$
 $35a + 20b \le 1190$
 $a,b \ge 0$

- Introduce variables *a* and *b* that define how much ale and beer to produce.
- Choose the variables in such a way that the objective function (profit) is maximized.
- Make sure that no constraints (due to limited supply) are violated.

- Introduce variables *a* and *b* that define how much ale and beer to produce.
- Choose the variables in such a way that the objective function (profit) is maximized.
- Make sure that no constraints (due to limited supply) are violated.

- Introduce variables *a* and *b* that define how much ale and beer to produce.
- Choose the variables in such a way that the objective function (profit) is maximized.
- Make sure that no constraints (due to limited supply) are violated.

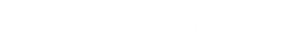
max
$$13a + 23b$$

s.t. $5a + 15b \le 480$
 $4a + 4b \le 160$
 $35a + 20b \le 1190$
 $a, b \ge 0$

- input: numbers a_{ij} , c_j , b_i
- \triangleright output: numbers x_1
- ightharpoonup n = #decision variables, m =
- maximize linear objective function subject to linear (in)equalities



- input: numbers a_{ij} , c_j , b_i
- ightharpoonup output: numbers x_j
- ightharpoonup n = #decision variables, m = #constraints
- maximize linear objective function subject to linear (in)equalities



- input: numbers a_{ij} , c_j , b_i
- \triangleright output: numbers x_i
- ightharpoonup n=#decision variables, m=#constraints
- maximize linear objective function subject to linear (in)equalities

- input: numbers a_{ij} , c_j , b_i
- \triangleright output: numbers x_i
- ightharpoonup n = #decision variables, m = #constraints
- maximize linear objective function subject to linear (in)equalities

$$\max \sum_{j=1}^{n} c_{j}x_{j}$$
s.t.
$$\sum_{j=1}^{n} a_{ij}x_{j} = b_{i} \quad 1 \leq i \leq m$$

$$x_{j} \geq 0 \quad 1 \leq j \leq n$$



- input: numbers a_{ij} , c_j , b_i
- output: numbers x_i
- ightharpoonup n=#decision variables, m=#constraints
- maximize linear objective function subject to linear (in)equalities

$$\max \sum_{j=1}^{n} c_{j}x_{j}$$
s.t.
$$\sum_{j=1}^{n} a_{ij}x_{j} = b_{i} \quad 1 \le i \le m$$

$$x_{j} \ge 0 \quad 1 \le j \le n$$

$$\begin{array}{rcl}
\max & c^T x \\
\text{s.t.} & Ax &= b \\
& x & \ge 0
\end{array}$$

- input: numbers a_{ij} , c_j , b_i
- \triangleright output: numbers x_i
- ightharpoonup n = #decision variables, m = #constraints
- maximize linear objective function subject to linear (in)equalities

$$\max \sum_{j=1}^{n} c_j x_j$$
s.t.
$$\sum_{j=1}^{n} a_{ij} x_j = b_i \quad 1 \le i \le m$$

$$x_j \ge 0 \quad 1 \le j \le n$$

$$\begin{array}{rcl}
\text{max} & c^T x \\
\text{s.t.} & Ax &= b \\
& x & \ge 0
\end{array}$$

Original LP

max
$$13a + 23b$$

s.t. $5a + 15b \le 480$
 $4a + 4b \le 160$
 $35a + 20b \le 1190$
 $a,b \ge 0$

Standard Form

Add a slack variable to every constraint

Original LP

max
$$13a + 23b$$

s.t. $5a + 15b \le 480$
 $4a + 4b \le 160$
 $35a + 20b \le 1190$
 $a, b \ge 0$

Standard Form

Add a slack variable to every constraint.

max
$$13a + 23b$$

s.t. $5a + 15b + s_c$ = 480
 $4a + 4b + s_h$ = 160
 $35a + 20b + s_h$ + $s_m = 1190$
 a , b , s_c , s_h , $s_m \ge 0$

There are different standard forms:

standard form

$$\max c^{T}x$$
s.t. $Ax = b$

$$x \ge 0$$

standard

maximization form

$$\begin{array}{rcl}
\text{max} & c^T x \\
\text{s.t.} & Ax & \leq b \\
& x & \geq 0
\end{array}$$

$\begin{array}{rcl} \min & c^T x \\ \text{s.t.} & Ax &= b \\ & x &\ge 0 \end{array}$

standard minimization fo

minimization form
$$\begin{array}{ccc}
\min & c^T x \\
\text{s.t.} & Ax \ge b \\
& x \ge 0
\end{array}$$

There are different standard forms:

standard form

$$\begin{array}{rcl}
\text{max} & c^T x \\
\text{s.t.} & Ax & = b \\
& x & \ge 0
\end{array}$$

standard

maximization form

$$\max c^T x$$
s.t. $Ax \le b$

$$x \ge 0$$

$\begin{array}{cccc} \min & c^T x \\ \text{s.t.} & Ax &= b \\ & x & \geq 0 \end{array}$

standard

minimization form

$$\begin{array}{rcl}
\min & c^T x \\
\text{s.t.} & Ax \ge b \\
& x \ge 0
\end{array}$$

There are different standard forms:

standard form

$$\max c^{T}x$$
s.t. $Ax = b$

$$x \ge 0$$

standard maximization form

$$\begin{array}{cccc}
\max & c^T x \\
\text{s.t.} & Ax & \leq & b \\
& x & \geq & 0
\end{array}$$

$$\begin{array}{rcl}
\min & c^T x \\
\text{s.t.} & Ax &= b \\
& x & \ge 0
\end{array}$$

standard minimization form

$$\begin{array}{rcl}
\min & c^T x \\
\text{s.t.} & Ax & \geq & b \\
& x & \geq & 0
\end{array}$$

There are different standard forms:

standard form

$$\max c^{T}x$$
s.t. $Ax = b$

$$x \ge 0$$

standard maximization form

$$\begin{array}{rcl}
\text{max} & c^T x \\
\text{s.t.} & Ax & \leq b \\
& x \geq 0
\end{array}$$

$$\begin{array}{rcl}
\min & c^T x \\
\text{s.t.} & Ax & = & b \\
& x & \ge & 0
\end{array}$$

standard minimization form

$$\begin{array}{cccc}
\min & c^T x \\
\text{s.t.} & Ax & \geq & b \\
& x & \geq & 0
\end{array}$$

It is easy to transform variants of LPs into (any) standard form:

It is easy to transform variants of LPs into (any) standard form:

less or equal to equality:

$$a - 3b + 5c \le 12 \implies a - 3b + 5c + s = 12$$
$$s \ge 0$$

It is easy to transform variants of LPs into (any) standard form:

less or equal to equality:

$$a - 3b + 5c \le 12 \implies a - 3b + 5c + s = 12$$
$$s \ge 0$$

greater or equal to equality:

min to max

It is easy to transform variants of LPs into (any) standard form:

less or equal to equality:

$$a - 3b + 5c \le 12 \implies a - 3b + 5c + s = 12$$
$$s \ge 0$$

greater or equal to equality:

$$a - 3b + 5c \ge 12 \implies a - 3b + 5c - s = 12$$

$$s \ge 0$$

min to max

It is easy to transform variants of LPs into (any) standard form:

less or equal to equality:

$$a-3b+5c \le 12 \implies a-3b+5c+s=12$$

 $s \ge 0$

greater or equal to equality:

$$a - 3b + 5c \ge 12 \implies a - 3b + 5c - s = 12$$
$$s \ge 0$$

min to max

It is easy to transform variants of LPs into (any) standard form:

less or equal to equality:

$$a - 3b + 5c \le 12 \implies a - 3b + 5c + s = 12$$
$$s \ge 0$$

greater or equal to equality:

$$a - 3b + 5c \ge 12 \implies a - 3b + 5c - s = 12$$
$$s \ge 0$$

min to max:

$$\min a - 3b + 5c \implies \max -a + 3b - 5c$$

It is easy to transform variants of LPs into (any) standard form:

less or equal to equality:

$$a - 3b + 5c \le 12 \implies a - 3b + 5c + s = 12$$
$$s \ge 0$$

greater or equal to equality:

$$a-3b+5c \ge 12 \implies a-3b+5c-s=12$$

 $s \ge 0$

min to max:

$$\min a - 3b + 5c \implies \max -a + 3b - 5c$$

It is easy to transform variants of LPs into (any) standard form:

equality to less or equal:

$$a - 3b + 5c = 12 \implies a - 3b + 5c \le 12$$

 $-a + 3b - 5c \le -12$

equality to greater or equal:

It is easy to transform variants of LPs into (any) standard form:

equality to less or equal:

$$a - 3b + 5c = 12 \implies a - 3b + 5c \le 12$$

 $-a + 3b - 5c \le -12$

equality to greater or equal:

It is easy to transform variants of LPs into (any) standard form:

equality to less or equal:

$$a - 3b + 5c = 12 \implies a - 3b + 5c \le 12$$

 $-a + 3b - 5c \le -12$

equality to greater or equal:

$$a-3b+5c = 12 \implies \frac{a-3b+5c \ge 12}{-a+3b-5c \ge -12}$$

It is easy to transform variants of LPs into (any) standard form:

equality to less or equal:

$$a - 3b + 5c = 12 \implies a - 3b + 5c \le 12$$

 $-a + 3b - 5c \le -12$

equality to greater or equal:

$$a - 3b + 5c = 12 \implies a - 3b + 5c \ge 12$$

 $-a + 3b - 5c \ge -12$

It is easy to transform variants of LPs into (any) standard form:

equality to less or equal:

$$a - 3b + 5c = 12 \implies a - 3b + 5c \le 12$$

 $-a + 3b - 5c \le -12$

equality to greater or equal:

$$a - 3b + 5c = 12 \implies a - 3b + 5c \ge 12$$

 $-a + 3b - 5c \ge -12$

$$x$$
 unrestricted $\Rightarrow x = x^+ - x^-, x^+ \ge 0, x^- \ge 0$

It is easy to transform variants of LPs into (any) standard form:

equality to less or equal:

$$a - 3b + 5c = 12 \implies a - 3b + 5c \le 12$$

 $-a + 3b - 5c \le -12$

equality to greater or equal:

$$a - 3b + 5c = 12 \implies a - 3b + 5c \ge 12$$

 $-a + 3b - 5c \ge -12$

$$x$$
 unrestricted $\Rightarrow x = x^+ - x^-, x^+ \ge 0, x^- \ge 0$

Observations:

- ▶ a linear program does not contain x^2 , $\cos(x)$, etc.
- transformations between standard forms can be done efficiently and only change the size of the LP by a small constant factor
- for the standard minimization or maximization LPs we could include the nonnegativity constraints into the set of ordinary constraints; this is of course not possible for the standard form

Observations:

- ▶ a linear program does not contain x^2 , $\cos(x)$, etc.
- transformations between standard forms can be done efficiently and only change the size of the LP by a small constant factor
- for the standard minimization or maximization LPs we could include the nonnegativity constraints into the set of ordinary constraints; this is of course not possible for the standard form

Observations:

- ▶ a linear program does not contain x^2 , cos(x), etc.
- transformations between standard forms can be done efficiently and only change the size of the LP by a small constant factor
- for the standard minimization or maximization LPs we could include the nonnegativity constraints into the set of ordinary constraints; this is of course not possible for the standard form

Definition 1 (Linear Programming Problem (LP))

Let $A \in \mathbb{Q}^{m \times n}$, $b \in \mathbb{Q}^m$, $c \in \mathbb{Q}^n$, $\alpha \in \mathbb{Q}$. Does there exist $x \in \mathbb{Q}^n$ s.t. Ax = b, $x \ge 0$, $c^Tx \ge \alpha$?

Ouestions:

- Is LP in NP?
- Is LP in co-NP?
- Is LP in P?

Input size:

Definition 1 (Linear Programming Problem (LP))

Let $A \in \mathbb{Q}^{m \times n}$, $b \in \mathbb{Q}^m$, $c \in \mathbb{Q}^n$, $\alpha \in \mathbb{Q}$. Does there exist $x \in \mathbb{Q}^n$ s.t. Ax = b, $x \ge 0$, $c^Tx \ge \alpha$?

Questions:

- Is I P in cn-NP2
- Is 1 P in P?

Input size:

Definition 1 (Linear Programming Problem (LP))

Let $A \in \mathbb{Q}^{m \times n}$, $b \in \mathbb{Q}^m$, $c \in \mathbb{Q}^n$, $\alpha \in \mathbb{Q}$. Does there exist $x \in \mathbb{Q}^n$ s.t. Ax = b, $x \ge 0$, $c^Tx \ge \alpha$?

Questions:

- ► Is LP in NP?
- ► Is LP in co-NP?
- ▶ Is I P in P?

Input size

Definition 1 (Linear Programming Problem (LP))

Let $A \in \mathbb{Q}^{m \times n}$, $b \in \mathbb{Q}^m$, $c \in \mathbb{Q}^n$, $\alpha \in \mathbb{Q}$. Does there exist $x \in \mathbb{Q}^n$ s.t. Ax = b, $x \ge 0$, $c^Tx \ge \alpha$?

Questions:

- ► Is I P in NP?
- ► Is LP in co-NP?
- ▶ Is I P in P?

Input size

Definition 1 (Linear Programming Problem (LP))

Let $A \in \mathbb{Q}^{m \times n}$, $b \in \mathbb{Q}^m$, $c \in \mathbb{Q}^n$, $\alpha \in \mathbb{Q}$. Does there exist $x \in \mathbb{Q}^n$ s.t. Ax = b, $x \ge 0$, $c^Tx \ge \alpha$?

Questions:

- ► Is I P in NP?
- ► Is LP in co-NP?
- Is I P in P?

Input size

Definition 1 (Linear Programming Problem (LP))

Let $A \in \mathbb{Q}^{m \times n}$, $b \in \mathbb{Q}^m$, $c \in \mathbb{Q}^n$, $\alpha \in \mathbb{Q}$. Does there exist $x \in \mathbb{Q}^n$ s.t. Ax = b, $x \ge 0$, $c^Tx \ge \alpha$?

Questions:

- ► Is LP in NP?
- ► Is LP in co-NP?
- Is I P in P?

Input size:

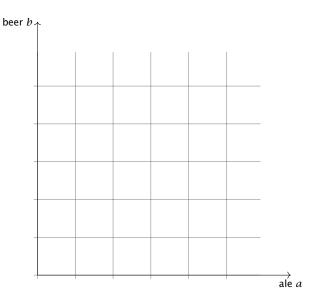
Definition 1 (Linear Programming Problem (LP))

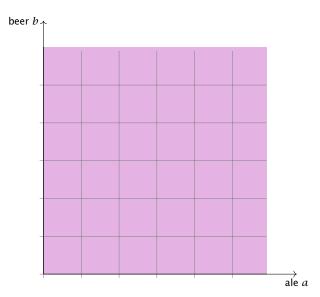
Let $A \in \mathbb{Q}^{m \times n}$, $b \in \mathbb{Q}^m$, $c \in \mathbb{Q}^n$, $\alpha \in \mathbb{Q}$. Does there exist $x \in \mathbb{Q}^n$ s.t. Ax = b, $x \ge 0$, $c^Tx \ge \alpha$?

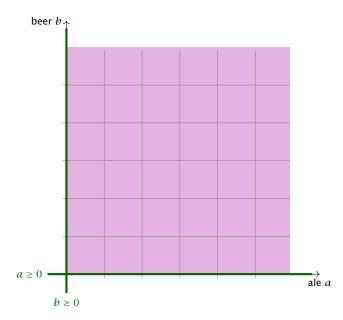
Questions:

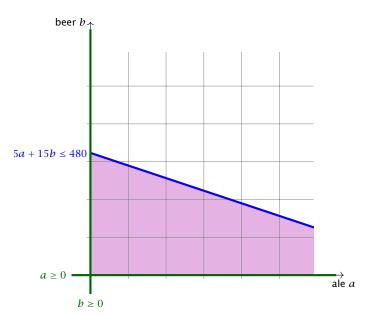
- ► Is LP in NP?
- ► Is LP in co-NP?
- Is I P in P?

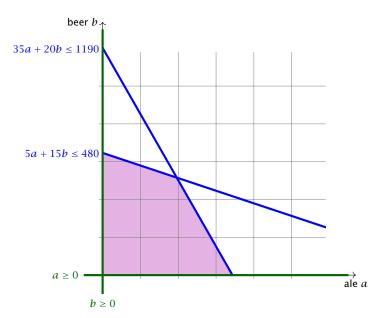
Input size:

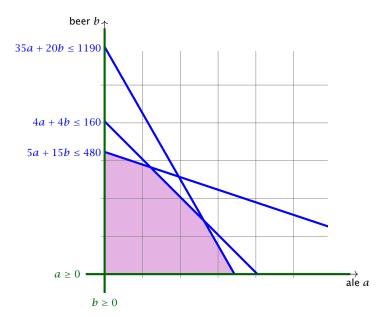


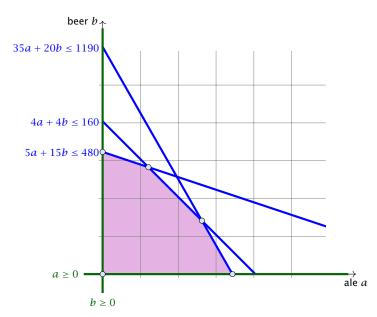


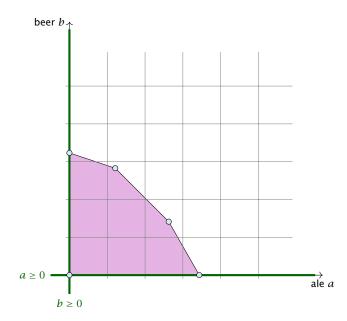


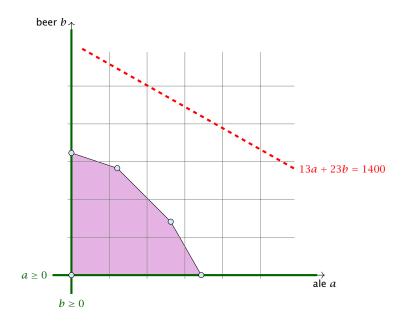


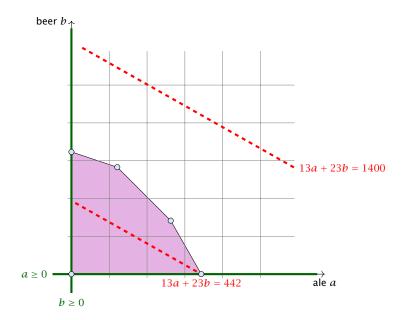


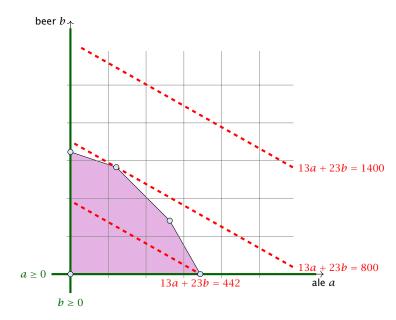


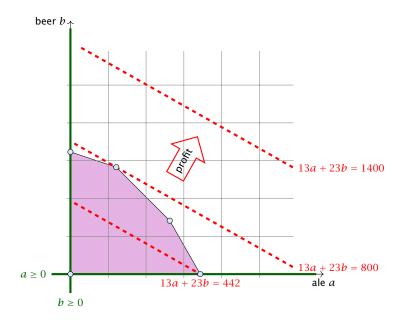


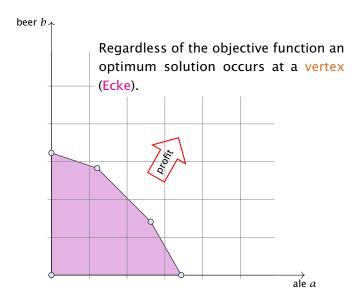












$$P = \{x \mid Ax = b, x \ge 0\}.$$

$$P = \{x \mid Ax = b, x \ge 0\}.$$

- P is called the feasible region (Lösungsraum) of the LP.
- ▶ A point $x \in P$ is called a feasible point (gültige Lösung)
- ▶ If $P \neq \emptyset$ then the LP is called feasible (erfüllbar). Otherwise
- An LP is bounded (beschränkt) if it is feasible and

$$P = \{x \mid Ax = b, x \ge 0\}.$$

- P is called the feasible region (Lösungsraum) of the LP.
- ▶ A point $x \in P$ is called a feasible point (gültige Lösung).
- ▶ If $P \neq \emptyset$ then the LP is called feasible (erfüllbar).
- ► An LP is bounded (beschränkt) if it is feasible and

$$P = \{x \mid Ax = b, x \ge 0\}.$$

- P is called the feasible region (Lösungsraum) of the LP.
- ▶ A point $x \in P$ is called a feasible point (gültige Lösung).
- ▶ If $P \neq \emptyset$ then the LP is called feasible (erfüllbar). Otherwise, it is called infeasible (unerfüllbar).
- An LP is bounded (beschränkt) if it is feasible and



$$P = \{x \mid Ax = b, x \ge 0\}.$$

- ▶ P is called the feasible region (Lösungsraum) of the LP.
- ▶ A point $x \in P$ is called a feasible point (gültige Lösung).
- If $P \neq \emptyset$ then the LP is called feasible (erfullbar). Otherwise, it is called infeasible (unerfullbar).
- ► An LP is bounded (beschränkt) if it is feasible and

Let for a Linear Program in standard form

$$P = \{x \mid Ax = b, x \ge 0\}.$$

- P is called the feasible region (Lösungsraum) of the LP.
- ▶ A point $x \in P$ is called a feasible point (gültige Lösung).
- If $P \neq \emptyset$ then the LP is called feasible (erfullbar). Otherwise, it is called infeasible (unerfullbar).
- An LP is bounded (beschränkt) if it is feasible and
 - $c^T x < \infty$ for all $x \in P$ (for maximization problems) $c^T x > -\infty$ for all $x \in P$ (for minimization problems)
 - $c^T x > -\infty$ for all $x \in P$ (for minimization problems)



Let for a Linear Program in standard form

$$P = \{x \mid Ax = b, x \ge 0\}.$$

- ▶ P is called the feasible region (Lösungsraum) of the LP.
- ▶ A point $x \in P$ is called a feasible point (gültige Lösung).
- If $P \neq \emptyset$ then the LP is called feasible (erfullbar). Otherwise, it is called infeasible (unerfullbar).
- An LP is bounded (beschränkt) if it is feasible and
 - $c^T x < \infty$ for all $x \in P$ (for maximization problems)
 - $c^{\perp}x > -\infty$ for all $x \in P$ (for minimization problems)



Let for a Linear Program in standard form

$$P = \{x \mid Ax = b, x \ge 0\}.$$

- ▶ *P* is called the feasible region (Lösungsraum) of the LP.
- ▶ A point $x \in P$ is called a feasible point (gültige Lösung).
- If $P \neq \emptyset$ then the LP is called feasible (erfullbar). Otherwise, it is called infeasible (unerfullbar).
- An LP is bounded (beschränkt) if it is feasible and
 - $ightharpoonup c^T x < \infty$ for all $x \in P$ (for maximization problems)
 - $c^T x > -\infty$ for all $x \in P$ (for minimization problems)

Given vectors/points $x_1, \ldots, x_k \in \mathbb{R}^n$, $\sum \lambda_i x_i$ is called

- ▶ linear combination if $\lambda_i \in \mathbb{R}$.
- ▶ affine combination if $\lambda_i \in \mathbb{R}$ and $\sum_i \lambda_i = 1$.
- **convex combination** if $\lambda_i \in \mathbb{R}$ and $\sum_i \lambda_i = 1$ and $\lambda_i \geq 0$.
- **conic combination** if $\lambda_i \in \mathbb{R}$ and $\lambda_i \geq 0$.

Note that a combination involves only finitely many vectors.

A set $X \subseteq \mathbb{R}^n$ is called

- a linear subspace if it is closed under linear combinations.
- an affine subspace if it is closed under affine combinations.
- convex if it is closed under convex combinations.
- a convex cone if it is closed under conic combinations.

Note that an affine subspace is **not** a vector space

Given a set $X \subseteq \mathbb{R}^n$.

- span(X) is the set of all linear combinations of X (linear hull, span)
- aff(X) is the set of all affine combinations of X (affine hull)
- conv(X) is the set of all convex combinations of X (convex hull)
- cone(X) is the set of all conic combinations of X (conic hull)

A function $f:\mathbb{R}^n\to\mathbb{R}$ is convex if for $x,y\in\mathbb{R}^n$ and $\lambda\in[0,1]$ we have

$$f(\lambda x + (1 - \lambda)y) \le \lambda f(x) + (1 - \lambda)f(y)$$

Lemma 6

If $P \subseteq \mathbb{R}^n$, and $f : \mathbb{R}^n \to \mathbb{R}$ convex then also

$$Q = \{x \in P \mid f(x) \le t\}$$

A function $f:\mathbb{R}^n\to\mathbb{R}$ is convex if for $x,y\in\mathbb{R}^n$ and $\lambda\in[0,1]$ we have

$$f(\lambda x + (1 - \lambda)y) \le \lambda f(x) + (1 - \lambda)f(y)$$

Lemma 6

If $P \subseteq \mathbb{R}^n$, and $f : \mathbb{R}^n \to \mathbb{R}$ convex then also

$$Q = \{x \in P \mid f(x) \le t\}$$

Dimensions

Definition 7

The dimension $\dim(A)$ of an affine subspace $A \subseteq \mathbb{R}^n$ is the dimension of the vector space $\{x - a \mid x \in A\}$, where $a \in A$.

Definition 8

The dimension $\dim(X)$ of a convex set $X \subseteq \mathbb{R}^n$ is the dimension of its affine hull $\operatorname{aff}(X)$.

A set $H \subseteq \mathbb{R}^n$ is a hyperplane if $H = \{x \mid a^T x = b\}$, for $a \neq 0$.

Definition 10

A set $H' \subseteq \mathbb{R}^n$ is a (closed) halfspace if $H = \{x \mid a^T x \le b\}$, for $a \ne 0$.

A set $H \subseteq \mathbb{R}^n$ is a hyperplane if $H = \{x \mid a^T x = b\}$, for $a \neq 0$.

Definition 10

A set $H' \subseteq \mathbb{R}^n$ is a (closed) halfspace if $H = \{x \mid a^Tx \leq b\}$, for $a \neq 0$.

Definition 11

A polytop is a set $P \subseteq \mathbb{R}^n$ that is the convex hull of a finite set of points, i.e., $P = \operatorname{conv}(X)$ where |X| = c.

Definition 12

A polyhedron is a set $P \subseteq \mathbb{R}^n$ that can be represented as the intersection of finitely many half-spaces

$$\{H(a_1, b_1), \dots, H(a_m, b_m)\}$$
, where

$$H(a_i,b_i) = \left\{ x \in \mathbb{R}^n \mid a_i x \le b_i \right\} .$$

Definition 13

A polyhedron P is bounded if there exists B s.t. $||x||_2 \le B$ for all $x \in P$.



Definition 12

A polyhedron is a set $P \subseteq \mathbb{R}^n$ that can be represented as the intersection of finitely many half-spaces

$$\{H(a_1, b_1), \dots, H(a_m, b_m)\}$$
, where

$$H(a_i,b_i) = \{x \in \mathbb{R}^n \mid a_i x \le b_i\} .$$

Definition 13

A polyhedron P is bounded if there exists B s.t. $||x||_2 \le B$ for all $x \in P$.



Theorem 14

P is a bounded polyhedron iff P is a polytop.

Let $P \subseteq \mathbb{R}^n$, $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$. The hyperplane

$$H(a,b) = \{x \in \mathbb{R}^n \mid a^T x = b\}$$

is a supporting hyperplane of P if $\max\{a^Tx \mid x \in P\} = b$.

Definition 16

Let $P \subseteq \mathbb{R}^n$. F is a face of P if F = P or $F = P \cap H$ for some supporting hyperplane H.

Definition 17

Let $P \subseteq \mathbb{R}^n$

- ightharpoonup a face v is a vertex of P if $\{v\}$ is a face of P.
- ightharpoonup a face e is an edge of P if e is a face and dim(e) = 1.
- ▶ a face *F* is a facet of *P* if *F* is a face and dim(F) = dim(P) 1

Let $P \subseteq \mathbb{R}^n$, $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$. The hyperplane

$$H(a,b) = \{x \in \mathbb{R}^n \mid a^T x = b\}$$

is a supporting hyperplane of P if $\max\{a^Tx \mid x \in P\} = b$.

Definition 16

Let $P \subseteq \mathbb{R}^n$. F is a face of P if F = P or $F = P \cap H$ for some supporting hyperplane H.

Definition 17

Let $P \subseteq \mathbb{R}^n$

- ightharpoonup a face v is a vertex of P if $\{v\}$ is a face of P.
- ightharpoonup a face e is an edge of P if e is a face and $\dim(e) = 1$.
- ightharpoonup a face F is a face of P if F is a face and $\dim(F) = \dim(P) 1$.

Let $P \subseteq \mathbb{R}^n$, $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$. The hyperplane

$$H(a,b) = \{x \in \mathbb{R}^n \mid a^T x = b\}$$

is a supporting hyperplane of P if $\max\{a^Tx \mid x \in P\} = b$.

Definition 16

Let $P \subseteq \mathbb{R}^n$. F is a face of P if F = P or $F = P \cap H$ for some supporting hyperplane H.

Definition 17

Let $P \subseteq \mathbb{R}^n$.

- ▶ a face *v* is a vertex of *P* if {*v*} is a face of *P*.
- ▶ a face e is an edge of P if e is a face and dim(e) = 1.
- ▶ a face *F* is a facet of *P* if *F* is a face and dim(F) = dim(P) 1.

Equivalent definition for vertex:

Definition 18

Given polyhedron P. A point $x \in P$ is a vertex if $\exists c \in \mathbb{R}^n$ such that $c^T y < c^T x$, for all $y \in P$, $y \neq x$.

Definition 19

Given polyhedron P. A point $x \in P$ is an extreme point if $\nexists a, b \neq x$, $a, b \in P$, with $\lambda a + (1 - \lambda)b = x$ for $\lambda \in [0, 1]$.

Lemma 20

A vertex is also an extreme point.

Equivalent definition for vertex:

Definition 18

Given polyhedron P. A point $x \in P$ is a vertex if $\exists c \in \mathbb{R}^n$ such that $c^T y < c^T x$, for all $y \in P$, $y \neq x$.

Definition 19

Given polyhedron P. A point $x \in P$ is an extreme point if $\nexists a, b \neq x, a, b \in P$, with $\lambda a + (1 - \lambda)b = x$ for $\lambda \in [0, 1]$.

Lemma 20

A vertex is also an extreme point.

Observation

The feasible region of an LP is a Polyhedron.

Theorem 21

If there exists an optimal solution to an LP (in standard form) then there exists an optimum solution that is an extreme point.

Theorem 21

If there exists an optimal solution to an LP (in standard form) then there exists an optimum solution that is an extreme point.

- suppose x is optimal solution that is not extreme point
- ▶ there exists direction $d \neq 0$ such that $x \pm d \in P$
- ightharpoonup Ad = 0 because $A(x \pm d) = b$
- ▶ Wlog. assume $c^T d \ge 0$ (by taking either d or -d)
- Consider $x + \lambda d$, $\lambda > 0$

Theorem 21

If there exists an optimal solution to an LP (in standard form) then there exists an optimum solution that is an extreme point.

- \triangleright suppose x is optimal solution that is not extreme point
- there exists direction $d \neq 0$ such that $x \pm d \in P$
- ightharpoonup Ad = 0 because $A(x \pm d) = b$
- ▶ Wlog. assume $c^T d \ge 0$ (by taking either d or -d)
- Consider $x + \lambda d$, $\lambda > 0$

Theorem 21

If there exists an optimal solution to an LP (in standard form) then there exists an optimum solution that is an extreme point.

- \triangleright suppose x is optimal solution that is not extreme point
- ▶ there exists direction $d \neq 0$ such that $x \pm d \in P$
- Ad = 0 because $A(x \pm d) = b$
- ▶ Wlog. assume $c^T d \ge 0$ (by taking either d or -d)
- Consider $x + \lambda d$, $\lambda > 0$

Theorem 21

If there exists an optimal solution to an LP (in standard form) then there exists an optimum solution that is an extreme point.

- \triangleright suppose x is optimal solution that is not extreme point
- ▶ there exists direction $d \neq 0$ such that $x \pm d \in P$
- Ad = 0 because $A(x \pm d) = b$
- ▶ Wlog. assume $c^T d \ge 0$ (by taking either d or -d)
- Consider $x + \lambda d$, $\lambda > 0$

Theorem 21

If there exists an optimal solution to an LP (in standard form) then there exists an optimum solution that is an extreme point.

- \triangleright suppose x is optimal solution that is not extreme point
- ▶ there exists direction $d \neq 0$ such that $x \pm d \in P$
- Ad = 0 because $A(x \pm d) = b$
- ▶ Wlog. assume $c^T d \ge 0$ (by taking either d or -d)
- Consider $x + \lambda d$, $\lambda > 0$

Case 1.
$$[\exists j \text{ s.t. } d_j < 0]$$

Case 2.
$$[d_j \ge 0 \text{ for all } j \text{ and } c^T d > 0]$$

is feasible for all since size so and and

as here, chia e hiller as

Case 1.
$$[\exists j \text{ s.t. } d_j < 0]$$

Case 1. $[\exists j \text{ s.t. } d_j < 0]$

- increase λ to λ' until first component of $x + \lambda d$ hits 0
- $ightharpoonup x + \lambda' d$ is feasible. Since $A(x + \lambda' d) = b$ and $x + \lambda' d \ge 0$
- $x + \lambda' d$ has one more zero-component ($d_k = 0$ for $x_k = 0$ as $x \pm d \in P$)
- $c^T x' = c^T (x + \lambda' d) = c^T x + \lambda' c^T d \ge c^T x$

Case 2. $[d_j \ge 0 \text{ for all } j \text{ and } c^T d > 0]$

is leasible for all A = 0 since A = A and a and



Case 1. $[\exists j \text{ s.t. } d_j < 0]$

- increase λ to λ' until first component of $x + \lambda d$ hits 0
- $x + \lambda' d$ is feasible. Since $A(x + \lambda' d) = b$ and $x + \lambda' d \ge 0$
- $x + \lambda' d$ has one more zero-component ($d_k = 0$ for $x_k = 0$ as $x \pm d \in P$)
- $c^T x' = c^T (x + \lambda' d) = c^T x + \lambda' c^T d \ge c^T x$

Case 2. $[d_j \ge 0 \text{ for all } j \text{ and } c^T d > 0]$

is feasible for all $\lambda=0$ since $\lambda(\lambda+\lambda \lambda)=\lambda$ and



Case 1. $[\exists j \text{ s.t. } d_j < 0]$

- increase λ to λ' until first component of $x + \lambda d$ hits 0
- $x + \lambda' d$ is feasible. Since $A(x + \lambda' d) = b$ and $x + \lambda' d \ge 0$
- $x + \lambda' d$ has one more zero-component ($d_k = 0$ for $x_k = 0$ as $x \pm d \in P$)
- $c^T x' = c^T (x + \lambda' d) = c^T x + \lambda' c^T d \ge c^T x$



Case 1. $[\exists j \text{ s.t. } d_j < 0]$

- increase λ to λ' until first component of $x + \lambda d$ hits 0
- $\triangleright x + \lambda' d$ is feasible. Since $A(x + \lambda' d) = b$ and $x + \lambda' d \ge 0$
- \times $x + \lambda' d$ has one more zero-component ($d_k = 0$ for $x_k = 0$ as $x \pm d \in P$)
- $c^T x' = c^T (x + \lambda' d) = c^T x + \lambda' c^T d \ge c^T x$

Case 1. $[\exists j \text{ s.t. } d_j < 0]$

- increase λ to λ' until first component of $x + \lambda d$ hits 0
- $x + \lambda' d$ is feasible. Since $A(x + \lambda' d) = b$ and $x + \lambda' d \ge 0$
- ► $x + \lambda' d$ has one more zero-component ($d_k = 0$ for $x_k = 0$ as $x \pm d \in P$)
- $c^T x' = c^T (x + \lambda' d) = c^T x + \lambda' c^T d \ge c^T x$

Case 1. $[\exists j \text{ s.t. } d_i < 0]$

- increase λ to λ' until first component of $x + \lambda d$ hits 0
- $\blacktriangleright x + \lambda' d$ is feasible. Since $A(x + \lambda' d) = b$ and $x + \lambda' d \ge 0$
- $x + \lambda' d$ has one more zero-component ($d_k = 0$ for $x_k = 0$ as $x \pm d \in P$)
- $c^T x' = c^T (x + \lambda' d) = c^T x + \lambda' c^T d \ge c^T x$

- ► $x + \lambda d$ is feasible for all $\lambda \ge 0$ since $A(x + \lambda d) = b$ and $x + \lambda d \ge x \ge 0$
- ightharpoonup as $\lambda \to \infty$, $c^T(x + \lambda d) \to \infty$ as $c^T d > 0$

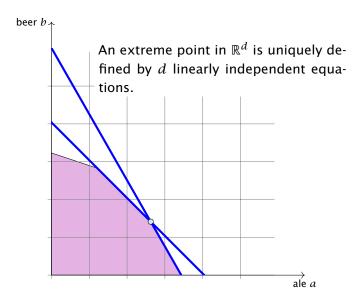


Case 1. $[\exists j \text{ s.t. } d_j < 0]$

- increase λ to λ' until first component of $x + \lambda d$ hits 0
- $\blacktriangleright x + \lambda' d$ is feasible. Since $A(x + \lambda' d) = b$ and $x + \lambda' d \ge 0$
- $x + \lambda' d$ has one more zero-component ($d_k = 0$ for $x_k = 0$ as $x \pm d \in P$)
- $c^T x' = c^T (x + \lambda' d) = c^T x + \lambda' c^T d \ge c^T x$

- ► $x + \lambda d$ is feasible for all $\lambda \ge 0$ since $A(x + \lambda d) = b$ and $x + \lambda d \ge x \ge 0$
- ▶ as $\lambda \to \infty$, $c^T(x + \lambda d) \to \infty$ as $c^T d > 0$

Algebraic View



Notation

Suppose $B \subseteq \{1 \dots n\}$ is a set of column-indices. Define A_B as the subset of columns of A indexed by B.

Theorem 22

Let $P = \{x \mid Ax = b, x \ge 0\}$. For $x \in P$, define $B = \{j \mid x_j > 0\}$. Then x is extreme point **iff** A_B has linearly independent columns.

Notation

Suppose $B \subseteq \{1 \dots n\}$ is a set of column-indices. Define A_B as the subset of columns of A indexed by B.

Theorem 22

Let $P = \{x \mid Ax = b, x \ge 0\}$. For $x \in P$, define $B = \{j \mid x_j > 0\}$.

Then x is extreme point **iff** A_B has linearly independent columns.

Let $P = \{x \mid Ax = b, x \ge 0\}$. For $x \in P$, define $B = \{j \mid x_j > 0\}$. Then x is extreme point **iff** A_B has linearly independent columns.

Proof (⇐)

Let $P = \{x \mid Ax = b, x \ge 0\}$. For $x \in P$, define $B = \{j \mid x_j > 0\}$. Then x is extreme point **iff** A_B has linearly independent columns.

- assume x is not extreme point
- ▶ there exists direction d s.t. $x \pm d \in P$
- Ad = 0 because $A(x \pm d) = b$
- $ightharpoonup A_{B'}$ has linearly dependent columns as Ad = 0
- $d_i = 0$ for all j with $x_i = 0$ as $x \pm d \ge 0$
- ▶ Hence, $B' \subseteq B$, $A_{B'}$ is sub-matrix of A_B

Let $P = \{x \mid Ax = b, x \ge 0\}$. For $x \in P$, define $B = \{j \mid x_j > 0\}$. Then x is extreme point **iff** A_B has linearly independent columns.

- assume x is not extreme point
- ▶ there exists direction d s.t. $x \pm d \in P$
- ightharpoonup Ad = 0 because $A(x \pm d) = b$
- $ightharpoonup A_{B'}$ has linearly dependent columns as Ad = 0
- $d_i = 0$ for all j with $x_i = 0$ as $x \pm d \ge 0$
- ► Hence, $B' \subseteq B$, $A_{B'}$ is sub-matrix of A_B

Let $P = \{x \mid Ax = b, x \ge 0\}$. For $x \in P$, define $B = \{j \mid x_j > 0\}$. Then x is extreme point **iff** A_B has linearly independent columns.

- assume x is not extreme point
- ▶ there exists direction d s.t. $x \pm d \in P$
- ightharpoonup Ad = 0 because $A(x \pm d) = b$
- ▶ define $B' = \{j \mid d_j \neq 0\}$
- $ightharpoonup A_{B'}$ has linearly dependent columns as Ad = 0
- $d_i = 0$ for all j with $x_i = 0$ as $x \pm d \ge 0$
- ▶ Hence, $B' \subseteq B$, $A_{B'}$ is sub-matrix of A_B

Let $P = \{x \mid Ax = b, x \ge 0\}$. For $x \in P$, define $B = \{j \mid x_j > 0\}$. Then x is extreme point **iff** A_B has linearly independent columns.

- assume x is not extreme point
- ▶ there exists direction d s.t. $x \pm d \in P$
- Ad = 0 because $A(x \pm d) = b$
- define $B' = \{j \mid d_j \neq 0\}$
- $ightharpoonup A_{B'}$ has linearly dependent columns as Ad=0
- $ightharpoonup d_j = 0$ for all j with $x_j = 0$ as $x \pm d \ge 0$
- ▶ Hence, $B' \subseteq B$, $A_{B'}$ is sub-matrix of A_B

Let $P = \{x \mid Ax = b, x \ge 0\}$. For $x \in P$, define $B = \{j \mid x_j > 0\}$. Then x is extreme point **iff** A_B has linearly independent columns.

- assume x is not extreme point
- ▶ there exists direction d s.t. $x \pm d \in P$
- ightharpoonup Ad = 0 because $A(x \pm d) = b$
- define $B' = \{j \mid d_j \neq 0\}$
- $ightharpoonup A_{B'}$ has linearly dependent columns as Ad=0
- $d_i = 0$ for all j with $x_i = 0$ as $x \pm d \ge 0$
- ▶ Hence, $B' \subseteq B$, $A_{B'}$ is sub-matrix of A_B

Let $P = \{x \mid Ax = b, x \ge 0\}$. For $x \in P$, define $B = \{j \mid x_j > 0\}$. Then x is extreme point **iff** A_B has linearly independent columns.

- assume x is not extreme point
- ▶ there exists direction d s.t. $x \pm d \in P$
- ightharpoonup Ad = 0 because $A(x \pm d) = b$
- define $B' = \{j \mid d_j \neq 0\}$
- $A_{B'}$ has linearly dependent columns as Ad = 0
- ▶ $d_j = 0$ for all j with $x_j = 0$ as $x \pm d \ge 0$
- ▶ Hence, $B' \subseteq B$, $A_{B'}$ is sub-matrix of A_B

Let $P = \{x \mid Ax = b, x \ge 0\}$. For $x \in P$, define $B = \{j \mid x_j > 0\}$. Then x is extreme point **iff** A_B has linearly independent columns.

- assume x is not extreme point
- ▶ there exists direction d s.t. $x \pm d \in P$
- ightharpoonup Ad = 0 because $A(x \pm d) = b$
- define $B' = \{j \mid d_j \neq 0\}$
- $ightharpoonup A_{B'}$ has linearly dependent columns as Ad=0
- ▶ $d_j = 0$ for all j with $x_j = 0$ as $x \pm d \ge 0$
- ▶ Hence, $B' \subseteq B$, $A_{B'}$ is sub-matrix of A_B

Let $P = \{x \mid Ax = b, x \ge 0\}$. For $x \in P$, define $B = \{j \mid x_j > 0\}$.

Then x is extreme point **iff** A_B has linearly independent columns.

Let $P = \{x \mid Ax = b, x \ge 0\}$. For $x \in P$, define $B = \{j \mid x_j > 0\}$. Then x is extreme point **iff** A_B has linearly independent columns.

- ightharpoonup assume A_B has linearly dependent columns
- ▶ there exists $d \neq 0$ such that $A_B d = 0$
- ightharpoonup extend d to \mathbb{R}^n by adding 0-components
- ightharpoonup now, Ad = 0 and $d_i = 0$ whenever $x_i = 0$
- ▶ for sufficiently small λ we have $x \pm \lambda d \in P$
- \triangleright hence, x is not extreme point

Let $P = \{x \mid Ax = b, x \ge 0\}$. For $x \in P$, define $B = \{j \mid x_j > 0\}$. Then x is extreme point **iff** A_B has linearly independent columns.

- ightharpoonup assume A_B has linearly dependent columns
- ▶ there exists $d \neq 0$ such that $A_B d = 0$
- ightharpoonup extend d to \mathbb{R}^n by adding 0-components
- ightharpoonup now, Ad = 0 and $d_i = 0$ whenever $x_i = 0$
- ▶ for sufficiently small λ we have $x \pm \lambda d \in P$
- \blacktriangleright hence, x is not extreme point

Let $P = \{x \mid Ax = b, x \ge 0\}$. For $x \in P$, define $B = \{j \mid x_j > 0\}$. Then x is extreme point **iff** A_B has linearly independent columns.

- ightharpoonup assume A_B has linearly dependent columns
- ▶ there exists $d \neq 0$ such that $A_B d = 0$
- extend d to \mathbb{R}^n by adding 0-components
- ▶ now, Ad = 0 and $d_j = 0$ whenever $x_j = 0$
- ▶ for sufficiently small λ we have $x \pm \lambda d \in P$
- \triangleright hence, x is not extreme point

Let $P = \{x \mid Ax = b, x \ge 0\}$. For $x \in P$, define $B = \{j \mid x_j > 0\}$. Then x is extreme point **iff** A_B has linearly independent columns.

- ightharpoonup assume A_B has linearly dependent columns
- there exists $d \neq 0$ such that $A_B d = 0$
- extend d to \mathbb{R}^n by adding 0-components
- ▶ now, Ad = 0 and $d_i = 0$ whenever $x_i = 0$
- ▶ for sufficiently small λ we have $x \pm \lambda d \in P$
- \blacktriangleright hence, x is not extreme point

Let $P = \{x \mid Ax = b, x \ge 0\}$. For $x \in P$, define $B = \{j \mid x_j > 0\}$. Then x is extreme point **iff** A_B has linearly independent columns.

- ightharpoonup assume A_B has linearly dependent columns
- there exists $d \neq 0$ such that $A_B d = 0$
- extend d to \mathbb{R}^n by adding 0-components
- ▶ now, Ad = 0 and $d_i = 0$ whenever $x_i = 0$
- for sufficiently small λ we have $x \pm \lambda d \in P$
- \triangleright hence, x is not extreme point

Let $P = \{x \mid Ax = b, x \ge 0\}$. For $x \in P$, define $B = \{j \mid x_j > 0\}$. Then x is extreme point **iff** A_B has linearly independent columns.

- ightharpoonup assume A_B has linearly dependent columns
- there exists $d \neq 0$ such that $A_B d = 0$
- extend d to \mathbb{R}^n by adding 0-components
- ▶ now, Ad = 0 and $d_i = 0$ whenever $x_i = 0$
- for sufficiently small λ we have $x \pm \lambda d \in P$
- hence, x is not extreme point

- ▶ then $c^T x = 0$ and $c^T y \le 0$ for $y \in P$
- ▶ assume $c^T y = 0$; then $y_j = 0$ for all $j \notin B$
- \blacktriangleright $b = Ay = A_By_B = Ax = A_Bx_B$ gives that $A_B(x_B y_B) = 0$;
- has linearly independent this means that $x_B = y_B$ since A_B has linearly independent columns
- we get y = x
- \blacktriangleright hence. x is a vertex of P

- ▶ define $c_j = \begin{cases} 0 & j \in B \\ -1 & j \notin B \end{cases}$
- ▶ then $c^T x = 0$ and $c^T y \le 0$ for $y \in P$
- ▶ assume $c^T y = 0$; then $y_i = 0$ for all $j \notin B$
- \blacktriangleright $b = Ay = A_By_B = Ax = A_Bx_B$ gives that $A_B(x_B y_B) = 0$;
- ▶ this means that $x_B = y_B$ since A_B has linearly independent columns
- we get y = x
- \blacktriangleright hence. x is a vertex of P

- ▶ then $c^T x = 0$ and $c^T y \le 0$ for $y \in P$
- ▶ assume $c^T y = 0$; then $y_j = 0$ for all $j \notin B$
- \blacktriangleright $b = Ay = A_By_B = Ax = A_Bx_B$ gives that $A_B(x_B y_B) = 0$;
- ▶ this means that $x_B = y_B$ since A_B has linearly independent columns
- we get y = x
- \blacktriangleright hence, x is a vertex of P

- ▶ then $c^T x = 0$ and $c^T y \le 0$ for $y \in P$
- ▶ assume $c^T y = 0$; then $y_j = 0$ for all $j \notin B$
- $b = Ay = A_By_B = Ax = A_Bx_B$ gives that $A_B(x_B y_B) = 0$;
- has linearly independent this means that $x_B = y_B$ since A_B has linearly independent columns
- we get y = x
- hence. x is a vertex of P

- $\bullet \ \, \mathsf{define} \,\, c_j = \left\{ \begin{array}{ll} 0 & j \in B \\ -1 & j \notin B \end{array} \right.$
- ▶ then $c^T x = 0$ and $c^T y \le 0$ for $y \in P$
- ▶ assume $c^T y = 0$; then $y_j = 0$ for all $j \notin B$
- $b = Ay = A_By_B = Ax = A_Bx_B$ gives that $A_B(x_B y_B) = 0$;
- ▶ this means that $x_B = y_B$ since A_B has linearly independent columns
- we get y = x
- \blacktriangleright hence, x is a vertex of P

- ▶ then $c^T x = 0$ and $c^T y \le 0$ for $y \in P$
- ▶ assume $c^T y = 0$; then $y_j = 0$ for all $j \notin B$
- $b = Ay = A_By_B = Ax = A_Bx_B$ gives that $A_B(x_B y_B) = 0$;
- ▶ this means that $x_B = y_B$ since A_B has linearly independent columns
- we get y = x
- \blacktriangleright hence, x is a vertex of P

- $\bullet \ \, \mathsf{define} \,\, c_j = \left\{ \begin{array}{ll} 0 & j \in B \\ -1 & j \notin B \end{array} \right.$
- ▶ then $c^T x = 0$ and $c^T y \le 0$ for $y \in P$
- ▶ assume $c^T y = 0$; then $y_j = 0$ for all $j \notin B$
- $b = Ay = A_By_B = Ax = A_Bx_B$ gives that $A_B(x_B y_B) = 0$;
- ▶ this means that $x_B = y_B$ since A_B has linearly independent columns
- we get y = x
- hence, x is a vertex of P

- ightharpoonup assume that $\operatorname{rank}(A) < m$
- ▶ assume wlog. that the first row A_1 lies in the span of the other rows A_2, \ldots, A_m ; this means

- C1 if now $b_1 = \sum_{i=2}^m \lambda_i \cdot b_i$ then for all with $b_1 = \sum_{i=2}^m \lambda_i \cdot b_i$ we also
- C2 if $b_1 \neq \sum_{i=2}^m \lambda_i \cdot b_i$ then the LP is infeasible, since for all x that fulfill constraints A_2, \ldots, A_m we have

- ightharpoonup assume that rank(A) < m
- ▶ assume wlog. that the first row A_1 lies in the span of the other rows A_2, \ldots, A_m ; this means

$$A_1 = \sum_{i=2}^m \lambda_i \cdot A_i$$
, for suitable λ_i

- C1 if now $b_1 = \sum_{i=2}^m \lambda_i \cdot b_i$ then for all a with
- C2 if $b_1 \neq \sum_{i=2}^m \lambda_i \cdot b_i$ then the LP is infeasible, since for all x that fulfill constraints A_2, \ldots, A_m we have

- ightharpoonup assume that rank(A) < m
- ▶ assume wlog. that the first row A_1 lies in the span of the other rows A_2, \ldots, A_m ; this means

$$A_1 = \sum_{i=2}^m \lambda_i \cdot A_i$$
, for suitable λ_i

- C1 if now $b_1 = \sum_{i=2}^m \lambda_i \cdot b_i$ then
- C2 if $b_1 \neq \sum_{i=2}^m \lambda_i \cdot b_i$ then the LP is infeasible, since for all x that fulfill constraints A_2, \ldots, A_m we have

- ightharpoonup assume that rank(A) < m
- ▶ assume wlog. that the first row A_1 lies in the span of the other rows A_2, \ldots, A_m ; this means

$$A_1 = \sum_{i=2}^m \lambda_i \cdot A_i$$
, for suitable λ_i

- C1 if now $b_1 = \sum_{i=2}^{m} \lambda_i \cdot b_i$ then for all x with $A_i x = b_i$ we also have $A_1 x = b_1$; hence the first constraint is superfluous
- C2 if $b_1 \neq \sum_{i=2}^m \lambda_i \cdot b_i$ then the LP is infeasible, since for all x that fulfill constraints A_2, \ldots, A_m we have

- ightharpoonup assume that rank(A) < m
- ▶ assume wlog. that the first row A_1 lies in the span of the other rows A_2, \ldots, A_m ; this means

$$A_1 = \sum_{i=2}^m \lambda_i \cdot A_i$$
, for suitable λ_i

- C1 if now $b_1 = \sum_{i=2}^m \lambda_i \cdot b_i$ then for all x with $A_i x = b_i$ we also have $A_1 x = b_1$; hence the first constraint is superfluous
- C2 if $b_1 \neq \sum_{i=2}^m \lambda_i \cdot b_i$ then the LP is infeasible, since for all x that fulfill constraints A_2, \ldots, A_m we have

- ightharpoonup assume that rank(A) < m
- ▶ assume wlog. that the first row A_1 lies in the span of the other rows A_2, \ldots, A_m ; this means

$$A_1 = \sum_{i=2}^m \lambda_i \cdot A_i$$
, for suitable λ_i

- C1 if now $b_1 = \sum_{i=2}^m \lambda_i \cdot b_i$ then for all x with $A_i x = b_i$ we also have $A_1 x = b_1$; hence the first constraint is superfluous
- C2 if $b_1 \neq \sum_{i=2}^m \lambda_i \cdot b_i$ then the LP is infeasible, since for all x that fulfill constraints A_2, \ldots, A_m we have

$$A_1 x = \sum_{i=2}^m \lambda_i \cdot A_i x = \sum_{i=2}^m \lambda_i \cdot b_i \neq b_1$$

- ightharpoonup assume that rank(A) < m
- ▶ assume wlog. that the first row A_1 lies in the span of the other rows A_2, \ldots, A_m ; this means

$$A_1 = \sum_{i=2}^m \lambda_i \cdot A_i$$
, for suitable λ_i

- C1 if now $b_1 = \sum_{i=2}^m \lambda_i \cdot b_i$ then for all x with $A_i x = b_i$ we also have $A_1 x = b_1$; hence the first constraint is superfluous
- C2 if $b_1 \neq \sum_{i=2}^m \lambda_i \cdot b_i$ then the LP is infeasible, since for all x that fulfill constraints A_2, \ldots, A_m we have

$$\mathbf{A_1x} = \sum_{i=2}^{m} \lambda_i \cdot A_i \mathbf{x} = \sum_{i=2}^{m} \lambda_i \cdot b_i \neq b_1$$

- ightharpoonup assume that rank(A) < m
- ▶ assume wlog. that the first row A_1 lies in the span of the other rows A_2, \ldots, A_m ; this means

$$A_1 = \sum_{i=2}^m \lambda_i \cdot A_i$$
, for suitable λ_i

- C1 if now $b_1 = \sum_{i=2}^m \lambda_i \cdot b_i$ then for all x with $A_i x = b_i$ we also have $A_1 x = b_1$; hence the first constraint is superfluous
- C2 if $b_1 \neq \sum_{i=2}^m \lambda_i \cdot b_i$ then the LP is infeasible, since for all x that fulfill constraints A_2, \ldots, A_m we have

$$A_1x = \sum_{i=2}^m \lambda_i \cdot A_ix = \sum_{i=2}^m \lambda_i \cdot b_i \neq b_1$$

- ightharpoonup assume that rank(A) < m
- ▶ assume wlog. that the first row A_1 lies in the span of the other rows A_2, \ldots, A_m ; this means

$$A_1 = \sum_{i=2}^m \lambda_i \cdot A_i$$
, for suitable λ_i

- C1 if now $b_1 = \sum_{i=2}^m \lambda_i \cdot b_i$ then for all x with $A_i x = b_i$ we also have $A_1 x = b_1$; hence the first constraint is superfluous
- C2 if $b_1 \neq \sum_{i=2}^m \lambda_i \cdot b_i$ then the LP is infeasible, since for all x that fulfill constraints A_2, \ldots, A_m we have

$$A_1x = \sum_{i=2}^m \lambda_i \cdot A_ix = \sum_{i=2}^m \lambda_i \cdot b_i \neq b_1$$

- ▶ assume that rank(A) < m</p>
- ▶ assume wlog. that the first row A_1 lies in the span of the other rows A_2, \ldots, A_m ; this means

$$A_1 = \sum_{i=2}^m \lambda_i \cdot A_i$$
, for suitable λ_i

- C1 if now $b_1 = \sum_{i=2}^m \lambda_i \cdot b_i$ then for all x with $A_i x = b_i$ we also have $A_1 x = b_1$; hence the first constraint is superfluous
- C2 if $b_1 \neq \sum_{i=2}^m \lambda_i \cdot b_i$ then the LP is infeasible, since for all x that fulfill constraints A_2, \ldots, A_m we have

$$A_1x = \sum_{i=2}^m \lambda_i \cdot A_ix = \sum_{i=2}^m \lambda_i \cdot b_i \neq b_1$$

From now on we will always assume that the constraint matrix of a standard form LP has full row rank.

Theorem 24

Given $P = \{x \mid Ax = b, x \ge 0\}$. x is extreme point iff there exists $B \subseteq \{1, ..., n\}$ with |B| = m and

- $ightharpoonup A_B$ is non-singular
- $\mathbf{x}_N = 0$

where $N = \{1, \ldots, n\} \setminus B$.

Proof

Take $B = \{j \mid x_j > 0\}$ and augment with linearly independent columns until |B| = m; always possible since rank(A) = m.

Theorem 24

Given $P = \{x \mid Ax = b, x \ge 0\}$. x is extreme point iff there exists $B \subseteq \{1, ..., n\}$ with |B| = m and

- $ightharpoonup A_B$ is non-singular
- $\rightarrow x_N = 0$

where $N = \{1, \ldots, n\} \setminus B$.

Proof

Take $B = \{j \mid x_j > 0\}$ and augment with linearly independent columns until |B| = m; always possible since rank(A) = m.

 $x \in \mathbb{R}^n$ is called basic solution (Basislösung) if Ax = b and $\operatorname{rank}(A_J) = |J|$ where $J = \{j \mid x_j \neq 0\}$;

x is a basic **feasible** solution (gültige Basislösung) if in addition $x \ge 0$.

A basis (Basis) is an index set $B \subseteq \{1, ..., n\}$ with $\operatorname{rank}(A_B) = m$ and |B| = m.

 $x \in \mathbb{R}^n$ is called basic solution (Basislösung) if Ax = b and $\operatorname{rank}(A_J) = |J|$ where $J = \{j \mid x_j \neq 0\}$;

x is a basic **feasible** solution (gültige Basislösung) if in addition $x \ge 0$.

A basis (Basis) is an index set $B \subseteq \{1, ..., n\}$ with $\operatorname{rank}(A_B) = m$ and |B| = m.

 $x \in \mathbb{R}^n$ is called basic solution (Basislösung) if Ax = b and $\operatorname{rank}(A_J) = |J|$ where $J = \{j \mid x_j \neq 0\}$;

x is a basic **feasible** solution (gültige Basislösung) if in addition $x \ge 0$.

A basis (Basis) is an index set $B \subseteq \{1, ..., n\}$ with $\mathrm{rank}(A_B) = m$ and |B| = m.

 $x \in \mathbb{R}^n$ is called basic solution (Basislösung) if Ax = b and $\operatorname{rank}(A_J) = |J|$ where $J = \{j \mid x_j \neq 0\}$;

x is a basic **feasible** solution (gültige Basislösung) if in addition $x \ge 0$.

A basis (Basis) is an index set $B \subseteq \{1, ..., n\}$ with $\mathrm{rank}(A_B) = m$ and |B| = m.

 $x \in \mathbb{R}^n$ is called basic solution (Basislösung) if Ax = b and $\operatorname{rank}(A_J) = |J|$ where $J = \{j \mid x_j \neq 0\}$;

x is a basic **feasible** solution (gültige Basislösung) if in addition $x \ge 0$.

A basis (Basis) is an index set $B \subseteq \{1, ..., n\}$ with $\mathrm{rank}(A_B) = m$ and |B| = m.

A BFS fulfills the m equality constraints.

In addition, at least n-m of the x_i 's are zero. The corresponding non-negativity constraint is fulfilled with equality.

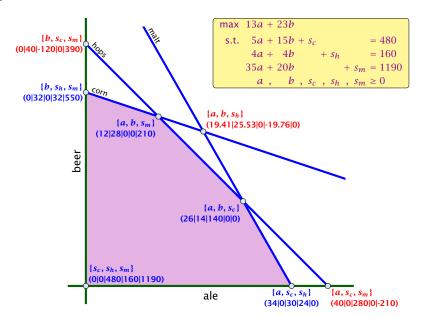
Fact:

In a BFS at least n constraints are fulfilled with equality.

Definition 25

For a general LP $(\max\{c^Tx \mid Ax \leq b\})$ with n variables a point x is a basic feasible solution if x is feasible and there exist n (linearly independent) constraints that are tight.

Algebraic View



Fundamental Questions

Linear Programming Problem (LP)

Let $A \in \mathbb{Q}^{m \times n}$, $b \in \mathbb{Q}^m$, $c \in \mathbb{Q}^n$, $\alpha \in \mathbb{Q}$. Does there exist $x \in \mathbb{Q}^n$ s.t. Ax = b, $x \ge 0$, $c^Tx \ge \alpha$?

Questions

- ► Is LP in NP? yes!
- ► Is LP in co-NP?
- ► Is LP in P?

Proof

Given a basis B we can compute the associated basis solution by calculating $A_B^{-1}b$ in polynomial time; then we can also compute the profit.

42/249

Fundamental Questions

Linear Programming Problem (LP)

Let $A \in \mathbb{Q}^{m \times n}$, $b \in \mathbb{Q}^m$, $c \in \mathbb{Q}^n$, $\alpha \in \mathbb{Q}$. Does there exist $x \in \mathbb{Q}^n$ s.t. Ax = b, $x \ge 0$, $c^Tx \ge \alpha$?

Questions:

- ► Is LP in NP? yes!
- ► Is LP in co-NP?
- ► Is LP in P?

Proof:

For Given a basis B we can compute the associated basis solution by calculating $A_B^{-1}b$ in polynomial time; then we can also compute the profit.

Observation

We can compute an optimal solution to a linear program in time $\mathcal{O}\left(\binom{n}{m}\cdot\operatorname{poly}(n,m)\right)$.

- there are only $\binom{n}{m}$ different bases.
- compute the profit of each of them and take the maximum

What happens if LP is unbounded?

Enumerating all basic feasible solutions (BFS), in order to find the optimum is slow.

Simplex Algorithm [George Dantzig 1947] Move from BFS to adjacent BFS, without decreasing objective function.

Two BFSs are called adjacent if the bases just differ in one variable.

Enumerating all basic feasible solutions (BFS), in order to find the optimum is slow.

Simplex Algorithm [George Dantzig 1947] Move from BFS to adjacent BFS, without decreasing objective function.

Two BFSs are called adjacent if the bases just differ in one variable.

max
$$13a + 23b$$

s.t. $5a + 15b + s_c = 480$
 $4a + 4b + s_h = 160$
 $35a + 20b + s_m = 1190$
 a , b , s_c , s_h , $s_m \ge 0$

basis = $\{s_c, s_h, s_m\}$ a = b = 0 Z = 0 $s_c = 480$ $s_h = 160$ $s_m = 1190$

$$\begin{array}{lllll} \max & 13a + 23b \\ \text{s.t.} & 5a + 15b + s_c & = 480 \\ & 4a + 4b & + s_h & = 160 \\ & 35a + 20b & + s_m = 1190 \\ & a & , & b & , s_c & , s_h & , s_m \geq 0 \end{array}$$

```
basis = \{s_c, s_h, s_m\}

a = b = 0

Z = 0

s_c = 480

s_h = 160

s_m = 1190
```

```
basis = \{s_c, s_h, s_m\}

a = b = 0

Z = 0

s_c = 480

s_h = 160

s_m = 1190
```

```
basis = \{s_c, s_h, s_m\}

a = b = 0

Z = 0

s_c = 480

s_h = 160

s_m = 1190
```

- choose variable to bring into the basis
- chosen variable should have positive coefficient in objective function
- apply min-ratio test to find out by how much the variable can be increased
- pivot on row found by min-ratio test
- the existing basis variable in this row leaves the basis

```
basis = \{s_c, s_h, s_m\}

a = b = 0

Z = 0

s_c = 480

s_h = 160

s_m = 1190
```

- choose variable to bring into the basis
- chosen variable should have positive coefficient in objective function
- apply min-ratio test to find out by how much the variable can be increased
- pivot on row found by min-ratio test
- the existing basis variable in this row leaves the basis

```
basis = \{s_c, s_h, s_m\}

a = b = 0

Z = 0

s_c = 480

s_h = 160

s_m = 1190
```

- choose variable to bring into the basis
- chosen variable should have positive coefficient in objective function
- apply min-ratio test to find out by how much the variable can be increased
- pivot on row found by min-ratio test
- the existing basis variable in this row leaves the basis

```
basis = \{s_c, s_h, s_m\}

a = b = 0

Z = 0

s_c = 480

s_h = 160

s_m = 1190
```

- choose variable to bring into the basis
- chosen variable should have positive coefficient in objective function
- apply min-ratio test to find out by how much the variable can be increased
- pivot on row found by min-ratio test
- the existing basis variable in this row leaves the basis

```
basis = \{s_c, s_h, s_m\}

a = b = 0

Z = 0

s_c = 480

s_h = 160

s_m = 1190
```

- choose variable to bring into the basis
- chosen variable should have positive coefficient in objective function
- apply min-ratio test to find out by how much the variable can be increased
- pivot on row found by min-ratio test
- the existing basis variable in this row leaves the basis

basis =
$$\{s_c, s_h, s_m\}$$

 $a = b = 0$
 $Z = 0$
 $s_c = 480$
 $s_h = 160$
 $s_m = 1190$

► Choose variable with coefficient > 0 as entering variable.

- Choose variable with coefficient > 0 as entering variable.
- If we keep a=0 and increase b from 0 to $\theta>0$ s.t. all constraints ($Ax=b, x\geq 0$) are still fulfilled the objective value Z will strictly increase.

basis =
$$\{s_c, s_h, s_m\}$$

 $a = b = 0$
 $Z = 0$
 $s_c = 480$
 $s_h = 160$
 $s_m = 1190$

- Choose variable with coefficient > 0 as entering variable.
- If we keep a=0 and increase b from 0 to $\theta>0$ s.t. all constraints ($Ax=b, x \ge 0$) are still fulfilled the objective value Z will strictly increase.
- For maintaining Ax = b we need e.g. to set $s_c = 480 15\theta$.

```
basis = \{s_c, s_h, s_m\}

a = b = 0

Z = 0

s_c = 480

s_h = 160

s_m = 1190
```

- Choose variable with coefficient > 0 as entering variable.
- If we keep a=0 and increase b from 0 to $\theta>0$ s.t. all constraints ($Ax=b,x\geq 0$) are still fulfilled the objective value Z will strictly increase.
- For maintaining Ax = b we need e.g. to set $s_c = 480 15\theta$.
- ► Choosing $\theta = \min\{480/15, 160/4, 1190/20\}$ ensures that in the new solution one current basic variable becomes 0, and no variable goes negative.

```
basis = \{s_c, s_h, s_m\}

a = b = 0

Z = 0

s_c = 480

s_h = 160

s_m = 1190
```

- Choose variable with coefficient > 0 as entering variable.
- If we keep a=0 and increase b from 0 to $\theta>0$ s.t. all constraints ($Ax=b, x \ge 0$) are still fulfilled the objective value Z will strictly increase.
- For maintaining Ax = b we need e.g. to set $s_c = 480 15\theta$.
- ► Choosing $\theta = \min\{480/15, 160/4, 1190/20\}$ ensures that in the new solution one current basic variable becomes 0, and no variable goes negative.
- The basic variable in the row that gives $min\{480/15, 160/4, 1190/20\}$ becomes the leaving variable.

Substitute $b = \frac{1}{15}(480 - 5a - s_c)$.

basis =
$$\{s_c, s_h, s_m\}$$

 $a = b = 0$
 $Z = 0$
 $s_c = 480$
 $s_h = 160$
 $s_m = 1190$

 ≥ 0

Substitute $b = \frac{1}{15}(480 - 5a - s_c)$.

a, b, s_c , s_h , s_m

$$\max Z$$

$$\frac{16}{3}a - \frac{23}{15}s_{c} - Z = -736$$

$$\frac{1}{3}a + b + \frac{1}{15}s_{c} = 32$$

$$\frac{8}{3}a - \frac{4}{15}s_{c} + s_{h} = 32$$

$$\frac{85}{3}a - \frac{4}{3}s_{c} + s_{m} = 550$$

$$a, b, s_{c}, s_{h}, s_{m} \ge 0$$

basis =
$$\{b, s_h, s_m\}$$

 $a = s_c = 0$
 $Z = 736$
 $b = 32$
 $s_h = 32$
 $s_m = 550$

$$\max Z$$

$$\frac{16}{3}a - \frac{23}{15}s_{c} - Z = -736$$

$$\frac{1}{3}a + b + \frac{1}{15}s_{c} = 32$$

$$\frac{8}{3}a - \frac{4}{15}s_{c} + s_{h} = 32$$

$$\frac{85}{3}a - \frac{4}{3}s_{c} + s_{m} = 550$$

$$a, b, s_{c}, s_{h}, s_{m} \ge 0$$

$$\max Z$$

$$\frac{16}{3}a - \frac{23}{15}s_{c} - Z = -736$$

$$\frac{1}{3}a + b + \frac{1}{15}s_{c} = 32$$

$$\frac{8}{3}a - \frac{4}{15}s_{c} + s_{h} = 32$$

$$\frac{85}{3}a - \frac{4}{3}s_{c} + s_{m} = 550$$

$$a, b, s_{c}, s_{h}, s_{m} \ge 0$$

Choose variable a to bring into basis.

$$\max Z$$

$$\frac{16}{3}a - \frac{23}{15}s_c - Z = -736$$

$$\frac{1}{3}a + b + \frac{1}{15}s_c - 32$$

$$\frac{8}{3}a - \frac{4}{15}s_c + s_h - 32$$

$$\frac{85}{3}a - \frac{4}{3}s_c + s_m - 550$$

$$a, b, s_c, s_h, s_m \ge 0$$

$$basis = \{b, s_h, s_m\}$$

$$a = s_c = 0$$

$$Z = 736$$

$$b = 32$$

$$s_h = 32$$

$$s_m = 550$$

Choose variable a to bring into basis. Computing $\min\{3 \cdot 32, \frac{3 \cdot 32}{8}, \frac{3 \cdot 550}{85}\}$ means pivot on line 2.

computing mm(5 × 52, 5 = 76, 5 = 5 / 65) means pivot on me 2.

Choose variable a to bring into basis.

Computing $\min\{3 \cdot 32, 3 \cdot 32/8, 3 \cdot 550/85\}$ means pivot on line 2. Substitute $a = \frac{3}{8}(32 + \frac{4}{15}s_c - s_h)$.

$$\max Z$$

$$\frac{16}{3}a - \frac{23}{15}s_{c} - Z = -736$$

$$\frac{1}{3}a + b + \frac{1}{15}s_{c} = 32$$

$$\frac{8}{3}a - \frac{4}{15}s_{c} + s_{h} = 32$$

$$\frac{85}{3}a - \frac{4}{3}s_{c} + s_{m} = 550$$

$$a, b, s_{c}, s_{h}, s_{m} \ge 0$$

basis = $\{b, s_h, s_m\}$ $a = s_c = 0$ Z = 736 b = 32 $s_h = 32$ $s_m = 550$

basis = $\{a, b, s_m\}$

 $s_c = s_h = 0$

Z = 800b = 28

a = 12

 $s_m = 210$

Choose variable *a* to bring into basis.

Computing $\min\{3 \cdot 32, 3 \cdot 32/8, 3 \cdot 550/85\}$ means pivot on line 2. Substitute $a = \frac{3}{8}(32 + \frac{4}{15}s_c - s_h)$.

$$\max Z - s_c - 2s_h - Z = -800$$

$$b + \frac{1}{10}s_c - \frac{1}{8}s_h = 28$$

$$a - \frac{1}{10}s_c + \frac{3}{8}s_h = 12$$

$$\frac{3}{2}s_c - \frac{85}{8}s_h + s_m = 210$$

$$a, b, s_c, s_h, s_m \ge 0$$

Pivoting stops when all coefficients in the objective function are non-positive.

Pivoting stops when all coefficients in the objective function are non-positive.

Pivoting stops when all coefficients in the objective function are non-positive.

- any feasible solution satisfies all equations in the tableaux
- ▶ in particular: $Z = 800 s_c 2s_h$, $s_c \ge 0$, $s_h \ge 0$
- ▶ hence optimum solution value is at most 800
- the current solution has value 800

Pivoting stops when all coefficients in the objective function are non-positive.

- any feasible solution satisfies all equations in the tableaux
- ▶ in particular: $Z = 800 s_c 2s_h$, $s_c \ge 0$, $s_h \ge 0$
- hence optimum solution value is at most 800
- the current solution has value 800

Pivoting stops when all coefficients in the objective function are non-positive.

- any feasible solution satisfies all equations in the tableaux
- ▶ in particular: $Z = 800 s_c 2s_h$, $s_c \ge 0$, $s_h \ge 0$
- hence optimum solution value is at most 800
- the current solution has value 800

Pivoting stops when all coefficients in the objective function are non-positive.

- any feasible solution satisfies all equations in the tableaux
- ▶ in particular: $Z = 800 s_c 2s_h$, $s_c \ge 0$, $s_h \ge 0$
- hence optimum solution value is at most 800
- the current solution has value 800

Let our linear program be

$$c_B^T x_B + c_N^T x_N = Z$$

$$A_B x_B + A_N x_N = b$$

$$x_B , x_N \ge 0$$

The simplex tableaux for basis *B* is

$$(c_N^T - c_B^T A_B^{-1} A_N) x_N = Z - c_B^T A_B^{-1} b$$

 $Ix_B + A_B^{-1} A_N x_N = A_B^{-1} b$
 $x_B , x_N \ge 0$

The BFS is given by $x_N = 0, x_B = A_B^{-1}b$.

If $(c_N^T - c_B^T A_B^{-1} A_N) \le 0$ we know that we have an optimum solution

51/249

Let our linear program be

$$c_B^T x_B + c_N^T x_N = Z$$

$$A_B x_B + A_N x_N = b$$

$$x_B , x_N \ge 0$$

The simplex tableaux for basis B is

$$(c_N^T - c_B^T A_B^{-1} A_N) x_N = Z - c_B^T A_B^{-1} b$$

 $Ix_B + A_B^{-1} A_N x_N = A_B^{-1} b$
 x_B , $x_N \ge 0$

The BFS is given by $x_N = 0, x_B = A_B^{-1}b$.



Let our linear program be

$$c_B^T x_B + c_N^T x_N = Z$$

$$A_B x_B + A_N x_N = b$$

$$x_B , x_N \ge 0$$

The simplex tableaux for basis B is

$$(c_N^T - c_B^T A_B^{-1} A_N) x_N = Z - c_B^T A_B^{-1} b$$

$$Ix_B + A_B^{-1} A_N x_N = A_B^{-1} b$$

$$x_B , x_N \ge 0$$

The BFS is given by $x_N = 0, x_B = A_B^{-1}b$.

Let our linear program be

$$c_B^T x_B + c_N^T x_N = Z$$

$$A_B x_B + A_N x_N = b$$

$$x_B , x_N \ge 0$$

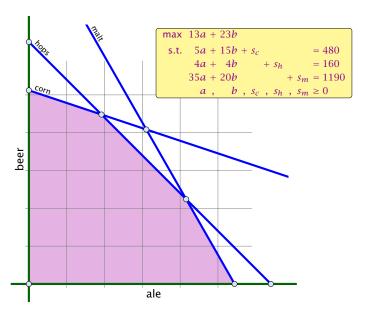
The simplex tableaux for basis B is

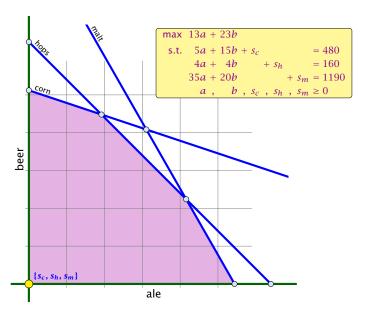
$$(c_N^T - c_B^T A_B^{-1} A_N) x_N = Z - c_B^T A_B^{-1} b$$

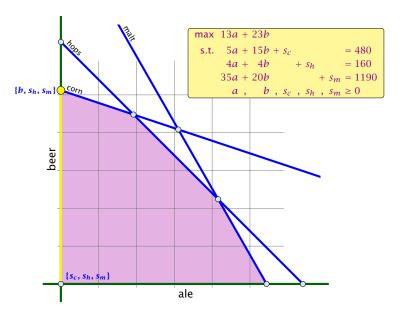
$$Ix_B + A_B^{-1} A_N x_N = A_B^{-1} b$$

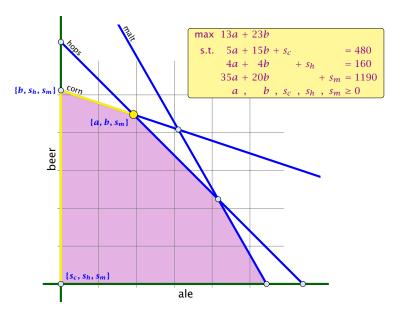
$$x_B , x_N \ge 0$$

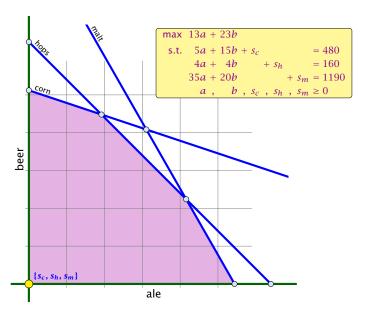
The BFS is given by $x_N = 0, x_B = A_B^{-1}b$.

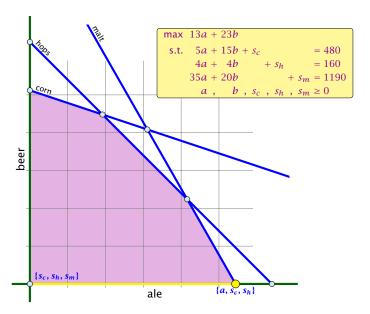


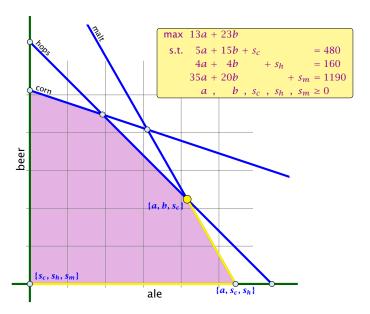


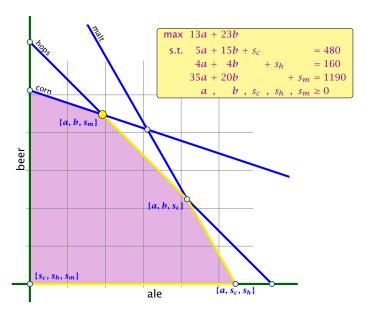












- Given basis B with BFS x^* .
- ► Choose index $j \notin B$ in order to increase x_i^* from 0 to $\theta > 0$.
 - Basis variables change to maintain feasibility
- ► Go from x^* to $x^* + \theta \cdot d$.

- (normalization)
- Aux + 6a h must hold. Hence
- Altogether: And An Add 11, which gives

- Given basis B with BFS x^* .
- ► Choose index $j \notin B$ in order to increase x_i^* from 0 to $\theta > 0$.
 - Other non-basis variables should stay at 0.
 - Basis variables change to maintain feasibility.
- ▶ Go from x^* to $x^* + \theta \cdot d$.

```
1 (normalization)
```

- No.2 0.4 0 must hold. Hence
- Altogether: And a A., Ad 0, which gives

- Given basis B with BFS x^* .
- ► Choose index $j \notin B$ in order to increase x_i^* from 0 to $\theta > 0$.
 - Other non-basis variables should stay at 0.
 - Basis variables change to maintain feasibility.
- ► Go from x^* to $x^* + \theta \cdot d$.



- Given basis B with BFS x^* .
- ► Choose index $j \notin B$ in order to increase x_j^* from 0 to $\theta > 0$.
 - Other non-basis variables should stay at 0.
 - Basis variables change to maintain feasibility.
- ► Go from x^* to $x^* + \theta \cdot d$.



- Given basis B with BFS x^* .
- ► Choose index $j \notin B$ in order to increase x_j^* from 0 to $\theta > 0$.
 - Other non-basis variables should stay at 0.
 - Basis variables change to maintain feasibility.
- ► Go from x^* to $x^* + \theta \cdot d$.

- Given basis B with BFS x^* .
- ► Choose index $j \notin B$ in order to increase x_j^* from 0 to $\theta > 0$.
 - Other non-basis variables should stay at 0.
 - Basis variables change to maintain feasibility.
- Go from x^* to $x^* + \theta \cdot d$.

- $ightharpoonup d_i = 1$ (normalization)
- $A(x^* + \theta d) = b$ must hold. Hence Ad = 0.
- Altogether: $A_B d_B + A_{*j} = Ad = 0$, which give: $d_B = -A_R^{-1} A_{*j}$.

- Given basis B with BFS x^* .
- ► Choose index $j \notin B$ in order to increase x_j^* from 0 to $\theta > 0$.
 - Other non-basis variables should stay at 0.
 - Basis variables change to maintain feasibility.
- Go from x^* to $x^* + \theta \cdot d$.

- $ightharpoonup d_j = 1$ (normalization)
- $A(x^* + \theta d) = b$ must hold. Hence Ad = 0.
- Altogether: $A_B d_B + A_{*j} = Ad = 0$, which gives $d_B = -A_B^{-1} A_{*j}$.



- Given basis B with BFS x^* .
- ► Choose index $j \notin B$ in order to increase x_i^* from 0 to $\theta > 0$.
 - Other non-basis variables should stay at 0.
 - Basis variables change to maintain feasibility.
- ► Go from x^* to $x^* + \theta \cdot d$.

- $ightharpoonup d_i = 1$ (normalization)
- $A(x^* + \theta d) = b$ must hold. Hence Ad = 0.
- Altogether: $A_B d_B + A_{*j} = Ad = 0$, which gives $d_B = -A_B^{-1} A_{*j}$.



- Given basis B with BFS x^* .
- ► Choose index $j \notin B$ in order to increase x_j^* from 0 to $\theta > 0$.
 - Other non-basis variables should stay at 0.
 - Basis variables change to maintain feasibility.
- ► Go from x^* to $x^* + \theta \cdot d$.

- $ightharpoonup d_j = 1$ (normalization)
- $A(x^* + \theta d) = b$ must hold. Hence Ad = 0.
- Altogether: $A_B d_B + A_{*j} = Ad = 0$, which gives $d_B = -A_B^{-1} A_{*j}$.

Definition 26 (j-th basis direction)

Let B be a basis, and let $j \notin B$. The vector d with $d_j = 1$ and $d_\ell = 0, \ell \notin B, \ell \neq j$ and $d_B = -A_B^{-1}A_{*j}$ is called the j-th basis direction for B.

Going from x^* to $x^* + \theta \cdot d$ the objective function changes by

$$\theta \cdot c^T d = \theta (c_j - c_B^T A_B^{-1} A_{*j})$$

Definition 26 (j-th basis direction)

Let B be a basis, and let $j \notin B$. The vector d with $d_j = 1$ and $d_\ell = 0, \ell \notin B, \ell \neq j$ and $d_B = -A_B^{-1}A_{*j}$ is called the j-th basis direction for B.

Going from x^* to $x^* + \theta \cdot d$ the objective function changes by

$$\theta \cdot c^T d = \theta (c_j - c_B^T A_B^{-1} A_{*j})$$

Definition 27 (Reduced Cost)

For a basis B the value

$$\tilde{c}_j = c_j - c_B^T A_B^{-1} A_{*j}$$

is called the reduced cost for variable x_j .

Note that this is defined for every j. If $j \in B$ then the above term is 0.

Let our linear program be

$$c_B^T x_B + c_N^T x_N = Z$$

$$A_B x_B + A_N x_N = b$$

$$x_B , x_N \ge 0$$

The simplex tableaux for basis B is

$$(c_N^T - c_B^T A_B^{-1} A_N) x_N = Z - c_B^T A_B^{-1} b$$

$$Ix_B + A_B^{-1} A_N x_N = A_B^{-1} b$$

$$x_N \ge 0$$

The BFS is given by $x_N = 0, x_B = A_B^{-1}b$.

Let our linear program be

$$c_B^T x_B + c_N^T x_N = Z$$

$$A_B x_B + A_N x_N = b$$

$$x_B , x_N \ge 0$$

The simplex tableaux for basis B is

$$(c_N^T - c_B^T A_B^{-1} A_N) x_N = Z - c_B^T A_B^{-1} b$$

$$Ix_B + A_B^{-1} A_N x_N = A_B^{-1} b$$

$$x_B , x_N \ge 0$$

The BFS is given by $x_N = 0, x_B = A_B^{-1}b$.

Let our linear program be

$$c_B^T x_B + c_N^T x_N = Z$$

$$A_B x_B + A_N x_N = b$$

$$x_B , x_N \ge 0$$

The simplex tableaux for basis B is

$$(c_N^T - c_B^T A_B^{-1} A_N) x_N = Z - c_B^T A_B^{-1} b$$

$$Ix_B + A_B^{-1} A_N x_N = A_B^{-1} b$$

$$x_B , x_N \ge 0$$

The BFS is given by $x_N = 0, x_B = A_B^{-1}b$.

Let our linear program be

$$c_B^T x_B + c_N^T x_N = Z$$

$$A_B x_B + A_N x_N = b$$

$$x_B , x_N \ge 0$$

The simplex tableaux for basis B is

$$(c_N^T - c_B^T A_B^{-1} A_N) x_N = Z - c_B^T A_B^{-1} b$$

 $Ix_B + A_B^{-1} A_N x_N = A_B^{-1} b$
 $x_B , x_N \ge 0$

The BFS is given by $x_N = 0$, $x_R = A_R^{-1}b$.

Questions:

Questions:

- What happens if the min ratio test fails to give us a value θ by which we can safely increase the entering variable?
- How do we find the initial basic feasible solution?
- Is there always a basis B such that

$$(c_N^T - c_B^T A_B^{-1} A_N) \le 0$$
 ?

- Then we can terminate because we know that the solution is optimal.
- If yes how do we make sure that we reach such a basis?

Questions:

- What happens if the min ratio test fails to give us a value θ by which we can safely increase the entering variable?
- How do we find the initial basic feasible solution?
- ▶ Is there always a basis *B* such that

$$(c_N^T - c_B^T A_B^{-1} A_N) \le 0$$
 ?

Then we can terminate because we know that the solution is optimal.

If yes how do we make sure that we reach such a basis?

Questions:

- ▶ What happens if the min ratio test fails to give us a value θ by which we can safely increase the entering variable?
- How do we find the initial basic feasible solution?
- Is there always a basis B such that

$$(c_N^T - c_B^T A_B^{-1} A_N) \le 0$$
 ?

Then we can terminate because we know that the solution is optimal.

▶ If yes how do we make sure that we reach such a basis?

Questions:

- What happens if the min ratio test fails to give us a value θ by which we can safely increase the entering variable?
- How do we find the initial basic feasible solution?
- Is there always a basis B such that

$$(c_N^T - c_B^T A_B^{-1} A_N) \le 0$$
 ?

Then we can terminate because we know that the solution is optimal.

If yes how do we make sure that we reach such a basis?

The min ratio test computes a value $\theta \ge 0$ such that after setting the entering variable to θ the leaving variable becomes 0 and all other variables stay non-negative.

For this, one computes b_i/A_{ie} for all constraints i and calculates the minimum positive value.

What does it mean that the ratio b_i/A_{ie} (and hence A_{ie}) is negative for a constraint?

This means that the corresponding basic variable will increase if we increase b. Hence, there is no danger of this basic variable becoming negative

The min ratio test computes a value $\theta \ge 0$ such that after setting the entering variable to θ the leaving variable becomes 0 and all other variables stay non-negative.

For this, one computes b_i/A_{ie} for all constraints i and calculates the minimum positive value.

What does it mean that the ratio b_i/A_{ie} (and hence A_{ie}) is negative for a constraint?

This means that the corresponding basic variable will increase if we increase b. Hence, there is no danger of this basic variable becoming negative

The min ratio test computes a value $\theta \ge 0$ such that after setting the entering variable to θ the leaving variable becomes 0 and all other variables stay non-negative.

For this, one computes b_i/A_{ie} for all constraints i and calculates the minimum positive value.

What does it mean that the ratio b_i/A_{ie} (and hence A_{ie}) is negative for a constraint?

This means that the corresponding basic variable will increase if we increase b. Hence, there is no danger of this basic variable becoming negative

The min ratio test computes a value $\theta \ge 0$ such that after setting the entering variable to θ the leaving variable becomes 0 and all other variables stay non-negative.

For this, one computes b_i/A_{ie} for all constraints i and calculates the minimum positive value.

What does it mean that the ratio b_i/A_{ie} (and hence A_{ie}) is negative for a constraint?

This means that the corresponding basic variable will increase if we increase b. Hence, there is no danger of this basic variable becoming negative

The min ratio test computes a value $\theta \ge 0$ such that after setting the entering variable to θ the leaving variable becomes 0 and all other variables stay non-negative.

For this, one computes b_i/A_{ie} for all constraints i and calculates the minimum positive value.

What does it mean that the ratio b_i/A_{ie} (and hence A_{ie}) is negative for a constraint?

This means that the corresponding basic variable will increase if we increase b. Hence, there is no danger of this basic variable becoming negative

The min ratio test computes a value $\theta \ge 0$ such that after setting the entering variable to θ the leaving variable becomes 0 and all other variables stay non-negative.

For this, one computes b_i/A_{ie} for all constraints i and calculates the minimum positive value.

What does it mean that the ratio b_i/A_{ie} (and hence A_{ie}) is negative for a constraint?

This means that the corresponding basic variable will increase if we increase b. Hence, there is no danger of this basic variable becoming negative

The objective function does not decrease during one iteration of the simplex-algorithm.

Does it always increase?

The objective function does not decrease during one iteration of the simplex-algorithm.

Does it always increase?

The objective function does not decrease during one iteration of the simplex-algorithm.

Does it always increase?

The objective function may not increase!

Because a variable x_{ℓ} with $\ell \in B$ is already 0.

The set of inequalities is degenerate (also the basis is degenerate).

Definition 28 (Degeneracy)

A BFS x^* is called degenerate if the set $J = \{j \mid x_j^* > 0\}$ fulfills |J| < m.

The objective function may not increase!

Because a variable x_{ℓ} with $\ell \in B$ is already 0.

The set of inequalities is degenerate (also the basis is degenerate).

Definition 28 (Degeneracy)

A BFS x^* is called degenerate if the set $J = \{j \mid x_j^* > 0\}$ fulfills |J| < m.

The objective function may not increase!

Because a variable x_{ℓ} with $\ell \in B$ is already 0.

The set of inequalities is degenerate (also the basis is degenerate).

Definition 28 (Degeneracy)

A BFS x^* is called degenerate if the set $J = \{j \mid x_j^* > 0\}$ fulfills |J| < m.

The objective function may not increase!

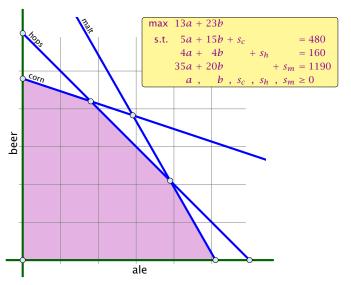
Because a variable x_{ℓ} with $\ell \in B$ is already 0.

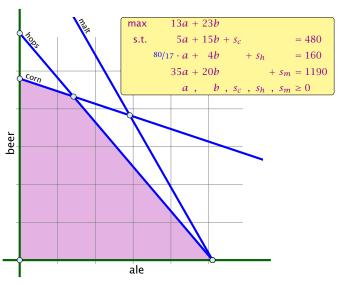
The set of inequalities is degenerate (also the basis is degenerate).

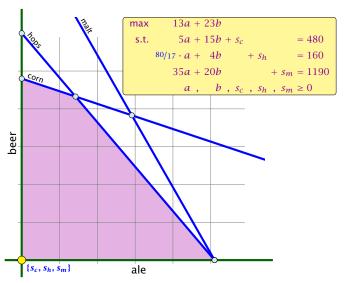
Definition 28 (Degeneracy)

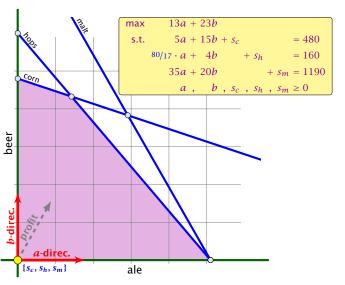
A BFS x^* is called degenerate if the set $J = \{j \mid x_j^* > 0\}$ fulfills |J| < m.

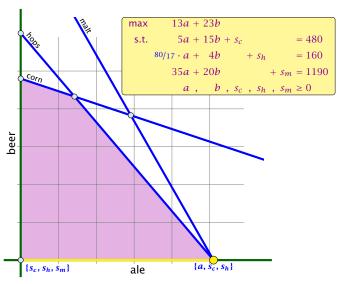
Non Degenerate Example

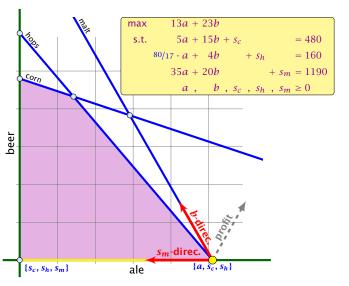


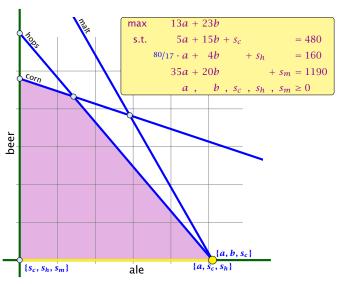


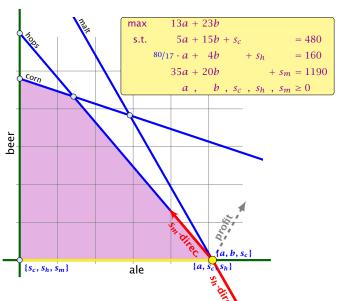


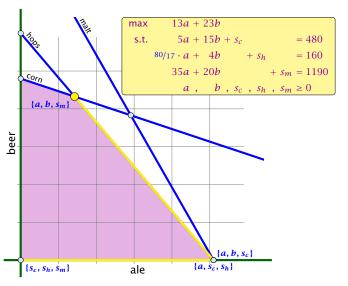


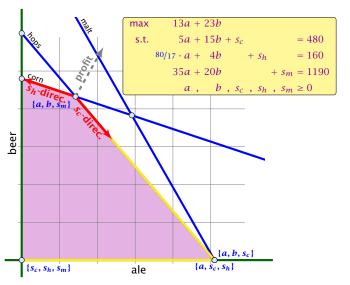












- We can choose a column e as an entering variable if $\tilde{c}_e > 0$ (\tilde{c}_e is reduced cost for x_e).
- ▶ The standard choice is the column that maximizes \tilde{c}_{ρ} ,
- ▶ If $A_{ie} \le 0$ for all $i \in \{1, ..., m\}$ then the maximum is not bounded.
- Otw. choose a leaving variable ℓ such that $b_{\ell}/A_{\ell e}$ is minimal among all variables i with $A_{ie} > 0$.
- ▶ If several variables have minimum $b_{\ell}/A_{\ell e}$ you reach a degenerate basis.
- ▶ Depending on the choice of ℓ it may happen that the algorithm runs into a cycle where it does not escape from a degenerate vertex.

- We can choose a column e as an entering variable if $\tilde{c}_e > 0$ (\tilde{c}_e is reduced cost for x_e).
- ▶ The standard choice is the column that maximizes \tilde{c}_e .
- ▶ If $A_{ie} \leq 0$ for all $i \in \{1, ..., m\}$ then the maximum is not bounded
- Otw. choose a leaving variable ℓ such that $b_{\ell}/A_{\ell e}$ is minimal among all variables i with $A_{ie} > 0$.
- If several variables have minimum $b_{\ell}/A_{\ell e}$ you reach a degenerate basis.
- ▶ Depending on the choice of ℓ it may happen that the algorithm runs into a cycle where it does not escape from a degenerate vertex.

- We can choose a column e as an entering variable if $\tilde{c}_e > 0$ (\tilde{c}_e is reduced cost for x_e).
- ▶ The standard choice is the column that maximizes \tilde{c}_e .
- ▶ If $A_{ie} \le 0$ for all $i \in \{1, ..., m\}$ then the maximum is not bounded.
- Otw. choose a leaving variable ℓ such that $b_{\ell}/A_{\ell e}$ is minimal among all variables i with $A_{ie} > 0$.
- If several variables have minimum $b_\ell/A_{\ell e}$ you reach a degenerate basis.
- ▶ Depending on the choice of ℓ it may happen that the algorithm runs into a cycle where it does not escape from a degenerate vertex.

- We can choose a column e as an entering variable if $\tilde{c}_e > 0$ (\tilde{c}_e is reduced cost for x_e).
- ▶ The standard choice is the column that maximizes \tilde{c}_e .
- ▶ If $A_{ie} \le 0$ for all $i \in \{1, ..., m\}$ then the maximum is not bounded.
- ▶ Otw. choose a leaving variable ℓ such that $b_{\ell}/A_{\ell e}$ is minimal among all variables i with $A_{ie} > 0$.
- If several variables have minimum $b_{\ell}/A_{\ell e}$ you reach a degenerate basis.
- ▶ Depending on the choice of ℓ it may happen that the algorithm runs into a cycle where it does not escape from a degenerate vertex.

- We can choose a column e as an entering variable if $\tilde{c}_e > 0$ (\tilde{c}_e is reduced cost for x_e).
- ▶ The standard choice is the column that maximizes \tilde{c}_e .
- ▶ If $A_{ie} \le 0$ for all $i \in \{1, ..., m\}$ then the maximum is not bounded.
- ▶ Otw. choose a leaving variable ℓ such that $b_{\ell}/A_{\ell e}$ is minimal among all variables i with $A_{ie} > 0$.
- If several variables have minimum $b_\ell/A_{\ell e}$ you reach a degenerate basis.
- ▶ Depending on the choice of ℓ it may happen that the algorithm runs into a cycle where it does not escape from a degenerate vertex.

- We can choose a column e as an entering variable if $\tilde{c}_e > 0$ (\tilde{c}_e is reduced cost for x_e).
- ▶ The standard choice is the column that maximizes \tilde{c}_e .
- ▶ If $A_{ie} \le 0$ for all $i \in \{1, ..., m\}$ then the maximum is not bounded.
- ▶ Otw. choose a leaving variable ℓ such that $b_{\ell}/A_{\ell e}$ is minimal among all variables i with $A_{ie} > 0$.
- If several variables have minimum $b_\ell/A_{\ell e}$ you reach a degenerate basis.
- ▶ Depending on the choice of ℓ it may happen that the algorithm runs into a cycle where it does not escape from a degenerate vertex.

What do we have so far?

Suppose we are given an initial feasible solution to an LP. If the LP is non-degenerate then Simplex will terminate.

Note that we either terminate because the min-ratio test fails and we can conclude that the LP is unbounded, or we terminate because the vector of reduced cost is non-positive. In the latter case we have an optimum solution.

How do we come up with an initial solution?

- $Ax \le b, x \ge 0$, and $b \ge 0$.
- ► The standard slack form for this problem is $Ax + Is = b, x \ge 0, s \ge 0$, where s denotes the vector of slack variables.
- ▶ Then s = b, x = 0 is a basic feasible solution (how?).
- We directly can start the simplex algorithm.

How do we find an initial basic feasible solution for an arbitrary problem?

- $Ax \le b, x \ge 0$, and $b \ge 0$.
- ► The standard slack form for this problem is $Ax + Is = b, x \ge 0, s \ge 0$, where s denotes the vector of slack variables.
- ▶ Then s = b, x = 0 is a basic feasible solution (how?).
- We directly can start the simplex algorithm.

- $Ax \le b, x \ge 0$, and $b \ge 0$.
- ► The standard slack form for this problem is $Ax + Is = b, x \ge 0, s \ge 0$, where s denotes the vector of slack variables.
- ▶ Then s = b, x = 0 is a basic feasible solution (how?).
- We directly can start the simplex algorithm.

- $Ax \le b, x \ge 0$, and $b \ge 0$.
- ► The standard slack form for this problem is $Ax + Is = b, x \ge 0, s \ge 0$, where s denotes the vector of slack variables.
- ▶ Then s = b, x = 0 is a basic feasible solution (how?).
- We directly can start the simplex algorithm.

- $Ax \le b, x \ge 0$, and $b \ge 0$.
- ► The standard slack form for this problem is $Ax + Is = b, x \ge 0, s \ge 0$, where s denotes the vector of slack variables.
- ▶ Then s = b, x = 0 is a basic feasible solution (how?).
- We directly can start the simplex algorithm.

```
Multiply all rows with by the by
```

Simplex.
$$x = 0$$
, $v = v$ is initial feasible.

If
$$\Sigma_1 = 0$$
 then the original problem is

- **1.** Multiply all rows with $b_i < 0$ by -1.
- 2. maximize $-\sum_i v_i$ s.t. Ax + Iv = b, $x \ge 0$, $v \ge 0$ using Simplex. x = 0, v = b is initial feasible.
- 3. If $\sum_i v_i > 0$ then the original problem is infeasible.
- **4.** Otw. you have $x \ge 0$ with Ax = b.
- **5.** From this you can get basic feasible solution.
- **6.** Now you can start the Simplex for the original problem



- **1.** Multiply all rows with $b_i < 0$ by -1.
- 2. maximize $-\sum_i v_i$ s.t. Ax + Iv = b, $x \ge 0$, $v \ge 0$ using Simplex. x = 0, v = b is initial feasible.
- 3. If $\sum_i v_i > 0$ then the original problem is infeasible.
- **4.** Otw. you have $x \ge 0$ with Ax = b.
- **5.** From this you can get basic feasible solution.
- 6. Now you can start the Simplex for the original problem.



- **1.** Multiply all rows with $b_i < 0$ by -1.
- 2. maximize $-\sum_i v_i$ s.t. Ax + Iv = b, $x \ge 0$, $v \ge 0$ using Simplex. x = 0, v = b is initial feasible.
- 3. If $\sum_i v_i > 0$ then the original problem is infeasible.
- **4.** Otw. you have $x \ge 0$ with Ax = b.
- **5.** From this you can get basic feasible solution.
- 6. Now you can start the Simplex for the original problem



- **1.** Multiply all rows with $b_i < 0$ by -1.
- 2. maximize $-\sum_i v_i$ s.t. Ax + Iv = b, $x \ge 0$, $v \ge 0$ using Simplex. x = 0, v = b is initial feasible.
- **3.** If $\sum_i v_i > 0$ then the original problem is infeasible.
- **4.** Otw. you have $x \ge 0$ with Ax = b.
- **5.** From this you can get basic feasible solution.
- 6. Now you can start the Simplex for the original problem.



- **1.** Multiply all rows with $b_i < 0$ by -1.
- 2. maximize $-\sum_i v_i$ s.t. Ax + Iv = b, $x \ge 0$, $v \ge 0$ using Simplex. x = 0, v = b is initial feasible.
- **3.** If $\sum_i v_i > 0$ then the original problem is infeasible.
- **4.** Otw. you have $x \ge 0$ with Ax = b.
- 5. From this you can get basic feasible solution.
- 6. Now you can start the Simplex for the original problem.

- **1.** Multiply all rows with $b_i < 0$ by -1.
- 2. maximize $-\sum_i v_i$ s.t. Ax + Iv = b, $x \ge 0$, $v \ge 0$ using Simplex. x = 0, v = b is initial feasible.
- **3.** If $\sum_i v_i > 0$ then the original problem is infeasible.
- **4.** Otw. you have $x \ge 0$ with Ax = b.
- 5. From this you can get basic feasible solution.
- 6. Now you can start the Simplex for the original problem.

Optimality

Lemma 29

Let B be a basis and x^* a BFS corresponding to basis B. $\tilde{c} \le 0$ implies that x^* is an optimum solution to the LP.

How do we get an upper bound to a maximization LP?

max
$$13a + 23b$$

s.t. $5a + 15b \le 480$
 $4a + 4b \le 160$
 $35a + 20b \le 1190$
 $a, b \ge 0$

Note that a lower bound is easy to derive. Every choice of $a, b \ge 0$ gives us a lower bound (e.g. a = 12, b = 28 gives us a lower bound of 800).

If you take a conic combination of the rows (multiply the i-th row with $y_i \geq 0$) such that $\sum_i y_i a_{ij} \geq c_j$ then $\sum_i y_i b_i$ will be an upper bound

How do we get an upper bound to a maximization LP?

max
$$13a + 23b$$

s.t. $5a + 15b \le 480$
 $4a + 4b \le 160$
 $35a + 20b \le 1190$
 $a, b \ge 0$

Note that a lower bound is easy to derive. Every choice of $a, b \ge 0$ gives us a lower bound (e.g. a = 12, b = 28 gives us a lower bound of 800).

If you take a conic combination of the rows (multiply the i-th row with $y_i \ge 0$) such that $\sum_i y_i a_{ij} \ge c_j$ then $\sum_i y_i b_i$ will be an upper bound.

How do we get an upper bound to a maximization LP?

max
$$13a + 23b$$

s.t. $5a + 15b \le 480$
 $4a + 4b \le 160$
 $35a + 20b \le 1190$
 $a, b \ge 0$

Note that a lower bound is easy to derive. Every choice of $a, b \ge 0$ gives us a lower bound (e.g. a = 12, b = 28 gives us a lower bound of 800).

If you take a conic combination of the rows (multiply the i-th row with $y_i \ge 0$) such that $\sum_i y_i a_{ij} \ge c_j$ then $\sum_i y_i b_i$ will be an upper bound.

Definition 30

Let $z = \max\{c^T x \mid Ax \le b, x \ge 0\}$ be a linear program P (called the primal linear program).

The linear program D defined by

$$w = \min\{b^T y \mid A^T y \ge c, y \ge 0\}$$

is called the dual problem.

Lemma 31

The dual of the dual problem is the primal problem.

Proof

Lemma 31

The dual of the dual problem is the primal problem.

Proof:

- $w = -\max\{-b^T y \mid -A^T y \le -c, y \ge 0\}$

Lemma 31

The dual of the dual problem is the primal problem.

Proof:

- $w = -\max\{-b^T y \mid -A^T y \le -c, y \ge 0\}$

Lemma 31

The dual of the dual problem is the primal problem.

Proof:

- $w = -\max\{-b^T y \mid -A^T y \le -c, y \ge 0\}$

- $z = -\min\{-c^T x \mid -Ax \ge -b, x \ge 0\}$
- $\triangleright z = \max\{c^T x \mid Ax \le b, x \ge 0\}$

Lemma 31

The dual of the dual problem is the primal problem.

Proof:

- $w = -\max\{-b^T y \mid -A^T y \le -c, y \ge 0\}$

- $z = -\min\{-c^T x \mid -Ax \ge -b, x \ge 0\}$
- $\triangleright z = \max\{c^T x \mid Ax \le b, x \ge 0\}$

Let
$$z = \max\{c^T x \mid Ax \le b, x \ge 0\}$$
 and $w = \min\{b^T y \mid A^T y \ge c, y \ge 0\}$ be a primal dual pair.

$$x$$
 is primal feasible iff $x \in \{x \mid Ax \le b, x \ge 0\}$

$$y$$
 is dual feasible, iff $y \in \{y \mid A^T y \ge c, y \ge 0\}$.

Theorem 32 (Weak Duality)

Let \hat{x} be primal feasible and let \hat{y} be dual feasible. Then

$$c^T\hat{x} \leq z \leq w \leq b^T\hat{y} \ .$$

Let $z = \max\{c^T x \mid Ax \le b, x \ge 0\}$ and $w = \min\{b^T y \mid A^T y \ge c, y \ge 0\}$ be a primal dual pair.

x is primal feasible iff $x \in \{x \mid Ax \le b, x \ge 0\}$

y is dual feasible, iff $y \in \{y \mid A^T y \ge c, y \ge 0\}$.

Theorem 32 (Weak Duality)

Let \hat{x} be primal feasible and let \hat{y} be dual feasible. Then

$$c^T \hat{x} \le z \le w \le b^T \hat{y}$$
.

$$A^T\hat{\mathcal{Y}} \geq c \Rightarrow \hat{x}^TA^T\hat{y} \geq \hat{x}^Tc\;(\hat{x} \geq 0)$$

$$A\hat{x} \le b \Rightarrow y^T A\hat{x} \le \hat{y}^T b \ (\hat{y} \ge 0)$$

This gives

$$c^T \hat{x} \le \hat{y}^T A \hat{x} \le b^T \hat{y}$$

Since, there exists primal feasible \hat{x} with $c^T\hat{x}=z$, and dual feasible \hat{y} with $b^T\hat{y}=w$ we get $z\leq w$.

$$A^T\hat{y} \geq c \Rightarrow \hat{x}^TA^T\hat{y} \geq \hat{x}^Tc \; (\hat{x} \geq 0)$$

$$A\hat{x} \le b \Rightarrow y^T A\hat{x} \le \hat{y}^T b \ (\hat{y} \ge 0)$$

This gives

$$c^T \hat{x} \le \hat{y}^T A \hat{x} \le b^T \hat{y} .$$

Since, there exists primal feasible \hat{x} with $c^T\hat{x} = z$, and dual feasible \hat{y} with $b^T\hat{y} = w$ we get $z \le w$.

$$A^T\hat{y} \geq c \Rightarrow \hat{x}^TA^T\hat{y} \geq \hat{x}^Tc \ (\hat{x} \geq 0)$$

$$A\hat{x} \le b \Rightarrow y^T A\hat{x} \le \hat{y}^T b \ (\hat{y} \ge 0)$$

This gives

$$c^T \hat{x} \le \hat{y}^T A \hat{x} \le b^T \hat{y} .$$

Since, there exists primal feasible \hat{x} with $c^T\hat{x}=z$, and dual feasible \hat{y} with $b^T\hat{y}=w$ we get $z\leq w$.

$$A^T\hat{y} \geq c \Rightarrow \hat{x}^TA^T\hat{y} \geq \hat{x}^Tc \ (\hat{x} \geq 0)$$

$$A\hat{\mathbf{x}} \leq \mathbf{b} \Rightarrow \mathbf{y}^T A \hat{\mathbf{x}} \leq \hat{\mathbf{y}}^T b \ (\hat{\mathbf{y}} \geq 0)$$

This gives

$$c^T\hat{x} \leq \hat{y}^T A \hat{x} \leq b^T \hat{y}$$

Since, there exists primal feasible \hat{x} with $c^T\hat{x} = z$, and dual feasible \hat{y} with $b^T\hat{y} = w$ we get $z \le w$.

$$A^T\hat{y} \geq c \Rightarrow \hat{x}^TA^T\hat{y} \geq \hat{x}^Tc \ (\hat{x} \geq 0)$$

$$A\hat{x} \leq b \Rightarrow y^T A \hat{x} \leq \hat{y}^T b \ (\hat{y} \geq 0)$$

This gives

$$c^T \hat{x} \le \hat{y}^T A \hat{x} \le b^T \hat{y}$$

Since, there exists primal feasible \hat{x} with $c^T\hat{x} = z$, and dual feasible \hat{y} with $b^T\hat{y} = w$ we get $z \le w$.

$$A^T\hat{y} \geq c \Rightarrow \hat{x}^TA^T\hat{y} \geq \hat{x}^Tc \ (\hat{x} \geq 0)$$

$$A\hat{x} \leq b \Rightarrow y^T A \hat{x} \leq \hat{y}^T b \; (\hat{y} \geq 0)$$

This gives

$$c^T \hat{x} \le \hat{y}^T A \hat{x} \le b^T \hat{y}$$

Since, there exists primal feasible \hat{x} with $c^T\hat{x} = z$, and dual feasible \hat{y} with $b^T\hat{y} = w$ we get $z \le w$.

$$A^T\hat{y} \geq c \Rightarrow \hat{x}^TA^T\hat{y} \geq \hat{x}^Tc \ (\hat{x} \geq 0)$$

$$A\hat{x} \le b \Rightarrow y^T A \hat{x} \le \hat{y}^T b \ (\hat{y} \ge 0)$$

This gives

$$c^T\hat{x} \leq \hat{y}^T A \hat{x} \leq b^T \hat{y} \ .$$

Since, there exists primal feasible \hat{x} with $c^T\hat{x} = z$, and dual feasible \hat{y} with $b^T\hat{y} = w$ we get $z \le w$.

$$A^T\hat{y} \geq c \Rightarrow \hat{x}^TA^T\hat{y} \geq \hat{x}^Tc \ (\hat{x} \geq 0)$$

$$A\hat{x} \le b \Rightarrow y^T A \hat{x} \le \hat{y}^T b \ (\hat{y} \ge 0)$$

This gives

$$c^T\hat{x} \leq \hat{y}^T A \hat{x} \leq b^T \hat{y} \ .$$

Since, there exists primal feasible \hat{x} with $c^T\hat{x}=z$, and dual feasible \hat{y} with $b^T\hat{y}=w$ we get $z\leq w$.

$$A^T\hat{y} \geq c \Rightarrow \hat{x}^TA^T\hat{y} \geq \hat{x}^Tc \ (\hat{x} \geq 0)$$

$$A\hat{x} \le b \Rightarrow y^T A\hat{x} \le \hat{y}^T b \ (\hat{y} \ge 0)$$

This gives

$$c^T\hat{x} \leq \hat{y}^T A \hat{x} \leq b^T \hat{y} \ .$$

Since, there exists primal feasible \hat{x} with $c^T\hat{x}=z$, and dual feasible \hat{y} with $b^T\hat{y}=w$ we get $z\leq w$.

5.2 Simplex and Duality

The following linear programs form a primal dual pair:

$$z = \max\{c^T x \mid Ax = b, x \ge 0\}$$
$$w = \min\{b^T y \mid A^T y \ge c\}$$

This means for computing the dual of a standard form LP, we do not have non-negativity constraints for the dual variables.

Proof

Primal:

$$\max\{c^T x \mid Ax = b, x \ge 0\}$$

Proof

Primal:

$$\max\{c^{T}x \mid Ax = b, x \ge 0\}$$

= \text{max}\{c^{T}x \ | Ax \le b, -Ax \le -b, x \ge 0\}

Proof

Primal:

$$\max\{c^{T}x \mid Ax = b, x \ge 0\}$$

$$= \max\{c^{T}x \mid Ax \le b, -Ax \le -b, x \ge 0\}$$

$$= \max\{c^{T}x \mid \begin{bmatrix} A \\ -A \end{bmatrix} x \le \begin{bmatrix} b \\ -b \end{bmatrix}, x \ge 0\}$$

Primal:

$$\max\{c^T x \mid Ax = b, x \ge 0\}$$

$$= \max\{c^T x \mid Ax \le b, -Ax \le -b, x \ge 0\}$$

$$= \max\{c^T x \mid \begin{bmatrix} A \\ -A \end{bmatrix} x \le \begin{bmatrix} b \\ -b \end{bmatrix}, x \ge 0\}$$

$$\min\{[b^T - b^T]y \mid [A^T - A^T]y \ge c, y \ge 0\}$$

Primal:

$$\max\{c^{T}x \mid Ax = b, x \ge 0\}$$

$$= \max\{c^{T}x \mid Ax \le b, -Ax \le -b, x \ge 0\}$$

$$= \max\{c^{T}x \mid \begin{bmatrix} A \\ -A \end{bmatrix} x \le \begin{bmatrix} b \\ -b \end{bmatrix}, x \ge 0\}$$

$$\min\{ \begin{bmatrix} b^T - b^T \end{bmatrix} y \mid \begin{bmatrix} A^T - A^T \end{bmatrix} y \ge c, y \ge 0 \}$$

$$= \min \left\{ \begin{bmatrix} b^T - b^T \end{bmatrix} \cdot \begin{bmatrix} y^+ \\ y^- \end{bmatrix} \mid \begin{bmatrix} A^T - A^T \end{bmatrix} \cdot \begin{bmatrix} y^+ \\ y^- \end{bmatrix} \ge c, y^- \ge 0, y^+ \ge 0 \right\}$$

Primal:

$$\max\{c^{T}x \mid Ax = b, x \ge 0\}$$

$$= \max\{c^{T}x \mid Ax \le b, -Ax \le -b, x \ge 0\}$$

$$= \max\{c^{T}x \mid \begin{bmatrix} A \\ -A \end{bmatrix} x \le \begin{bmatrix} b \\ -b \end{bmatrix}, x \ge 0\}$$

$$\min\{ \begin{bmatrix} b^T - b^T \end{bmatrix} y \mid \begin{bmatrix} A^T - A^T \end{bmatrix} y \ge c, y \ge 0 \}$$

$$= \min \left\{ \begin{bmatrix} b^T - b^T \end{bmatrix} \cdot \begin{bmatrix} y^+ \\ y^- \end{bmatrix} \mid \begin{bmatrix} A^T - A^T \end{bmatrix} \cdot \begin{bmatrix} y^+ \\ y^- \end{bmatrix} \ge c, y^- \ge 0, y^+ \ge 0 \right\}$$

$$= \min \left\{ b^T \cdot (y^+ - y^-) \mid A^T \cdot (y^+ - y^-) \ge c, y^- \ge 0, y^+ \ge 0 \right\}$$

Primal:

$$\max\{c^{T}x \mid Ax = b, x \ge 0\}$$

$$= \max\{c^{T}x \mid Ax \le b, -Ax \le -b, x \ge 0\}$$

$$= \max\{c^{T}x \mid \begin{bmatrix} A \\ -A \end{bmatrix} x \le \begin{bmatrix} b \\ -b \end{bmatrix}, x \ge 0\}$$

$$\begin{aligned} & \min\{ \left[b^T - b^T \right] y \mid \left[A^T - A^T \right] y \geq c, y \geq 0 \} \\ &= \min\left\{ \left[b^T - b^T \right] \cdot \begin{bmatrix} y^+ \\ y^- \end{bmatrix} \mid \left[A^T - A^T \right] \cdot \begin{bmatrix} y^+ \\ y^- \end{bmatrix} \geq c, y^- \geq 0, y^+ \geq 0 \right\} \\ &= \min\left\{ b^T \cdot (y^+ - y^-) \mid A^T \cdot (y^+ - y^-) \geq c, y^- \geq 0, y^+ \geq 0 \right\} \\ &= \min\left\{ b^T y' \mid A^T y' \geq c \right\} \end{aligned}$$

Suppose that we have a basic feasible solution with reduced cost

$$\tilde{c} = c^T - c_B^T A_B^{-1} A \leq 0$$

This is equivalent to $A^T(A_B^{-1})^Tc_B \ge c$

$$y^* = (A_B^{-1})^T c_B$$
 is solution to the dual $\min\{b^T y | A^T y \ge c\}$.

Suppose that we have a basic feasible solution with reduced cost

$$\tilde{c} = c^T - c_B^T A_B^{-1} A \leq 0$$

This is equivalent to $A^T(A_R^{-1})^T c_B \ge c$

 $y^* = (A_B^{-1})^T c_B$ is solution to the dual $\min\{b^T y | A^T y \ge c\}$.

Suppose that we have a basic feasible solution with reduced cost

$$\tilde{c} = c^T - c_B^T A_B^{-1} A \leq 0$$

This is equivalent to $A^T(A_B^{-1})^T c_B \ge c$

$$y^* = (A_B^{-1})^T c_B$$
 is solution to the dual $\min\{b^T y | A^T y \ge c\}$.

$$b^{T}y^{*} = (Ax^{*})^{T}y^{*} = (A_{B}x_{B}^{*})^{T}y^{*}$$
$$= (A_{B}x_{B}^{*})^{T}(A_{B}^{-1})^{T}c_{B} = (x_{B}^{*})^{T}A_{B}^{T}(A_{B}^{-1})^{T}c_{B}$$
$$= c^{T}x^{*}$$

Suppose that we have a basic feasible solution with reduced cost

$$\tilde{c} = c^T - c_B^T A_B^{-1} A \leq 0$$

This is equivalent to $A^T(A_B^{-1})^T c_B \ge c$

$$y^* = (A_B^{-1})^T c_B$$
 is solution to the dual $\min\{b^T y | A^T y \ge c\}$.

$$b^{T}y^{*} = (Ax^{*})^{T}y^{*} = (A_{B}x_{B}^{*})^{T}y^{*}$$
$$= (A_{B}x_{B}^{*})^{T}(A_{B}^{-1})^{T}c_{B} = (x_{B}^{*})^{T}A_{B}^{T}(A_{B}^{-1})^{T}c_{B}$$
$$= c^{T}x^{*}$$

Suppose that we have a basic feasible solution with reduced cost

$$\tilde{c} = c^T - c_B^T A_B^{-1} A \leq 0$$

This is equivalent to $A^T(A_B^{-1})^T c_B \ge c$

$$y^* = (A_B^{-1})^T c_B$$
 is solution to the dual $\min\{b^T y | A^T y \ge c\}$.

$$b^{T}y^{*} = (Ax^{*})^{T}y^{*} = (A_{B}x_{B}^{*})^{T}y^{*}$$
$$= (A_{B}x_{B}^{*})^{T}(A_{B}^{-1})^{T}c_{B} = (x_{B}^{*})^{T}A_{B}^{T}(A_{B}^{-1})^{T}c_{B}$$
$$= c^{T}x^{*}$$

Suppose that we have a basic feasible solution with reduced cost

$$\tilde{c} = c^T - c_B^T A_B^{-1} A \leq 0$$

This is equivalent to $A^T(A_B^{-1})^T c_B \ge c$

$$y^* = (A_B^{-1})^T c_B$$
 is solution to the dual $\min\{b^T y | A^T y \ge c\}$.

$$b^{T}y^{*} = (Ax^{*})^{T}y^{*} = (A_{B}x_{B}^{*})^{T}y^{*}$$
$$= (A_{B}x_{B}^{*})^{T}(A_{B}^{-1})^{T}c_{B} = (x_{B}^{*})^{T}A_{B}^{T}(A_{B}^{-1})^{T}c_{B}$$
$$= c^{T}x^{*}$$

Suppose that we have a basic feasible solution with reduced cost

$$\tilde{c} = c^T - c_B^T A_B^{-1} A \leq 0$$

This is equivalent to $A^T(A_B^{-1})^T c_B \ge c$

$$y^* = (A_B^{-1})^T c_B$$
 is solution to the dual $\min\{b^T y | A^T y \ge c\}$.

$$b^{T}y^{*} = (Ax^{*})^{T}y^{*} = (A_{B}x_{B}^{*})^{T}y^{*}$$
$$= (A_{B}x_{B}^{*})^{T}(A_{B}^{-1})^{T}c_{B} = (x_{B}^{*})^{T}A_{B}^{T}(A_{B}^{-1})^{T}c_{B}$$

Suppose that we have a basic feasible solution with reduced cost

$$\tilde{c} = c^T - c_B^T A_B^{-1} A \leq 0$$

This is equivalent to $A^T(A_B^{-1})^T c_B \ge c$

$$y^* = (A_B^{-1})^T c_B$$
 is solution to the dual $\min\{b^T y | A^T y \ge c\}$.

$$b^{T}y^{*} = (Ax^{*})^{T}y^{*} = (A_{B}x_{B}^{*})^{T}y^{*}$$
$$= (A_{B}x_{B}^{*})^{T}(A_{B}^{-1})^{T}c_{B} = (x_{B}^{*})^{T}A_{B}^{T}(A_{B}^{-1})^{T}c_{B}$$
$$= c^{T}x^{*}$$

Suppose that we have a basic feasible solution with reduced cost

$$\tilde{c} = c^T - c_B^T A_B^{-1} A \le 0$$

This is equivalent to $A^T(A_R^{-1})^T c_B \ge c$

$$y^* = (A_B^{-1})^T c_B$$
 is solution to the dual $\min\{b^T y | A^T y \ge c\}$.

$$b^{T}y^{*} = (Ax^{*})^{T}y^{*} = (A_{B}x_{B}^{*})^{T}y^{*}$$
$$= (A_{B}x_{B}^{*})^{T}(A_{B}^{-1})^{T}c_{B} = (x_{B}^{*})^{T}A_{B}^{T}(A_{B}^{-1})^{T}c_{B}$$
$$= c^{T}x^{*}$$

5.3 Strong Duality

$$P = \max\{c^T x \mid Ax \le b, x \ge 0\}$$

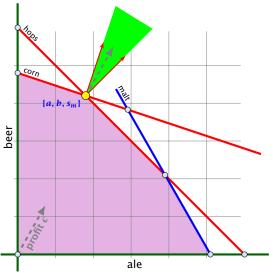
 n_A : number of variables, m_A : number of constraints

We can put the non-negativity constraints into A (which gives us unrestricted variables): $\bar{P} = \max\{c^T x \mid \bar{A}x \leq \bar{b}\}$

$$n_{\bar{A}}=n_A$$
, $m_{\bar{A}}=m_A+n_A$

Dual
$$D = \min\{\bar{b}^T y \mid \bar{A}^T y = c, y \ge 0\}.$$

5.3 Strong Duality



The profit vector \boldsymbol{c} lies in the cone generated by the normals for the hops and the corn constraint (the tight constraints).

Strong Duality

Theorem 33 (Strong Duality)

Let P and D be a primal dual pair of linear programs, and let z^* and w^* denote the optimal solution to P and D, respectively. Then

$$z^* = w^*$$

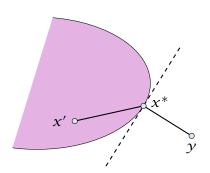
Lemma 34 (Weierstrass)

Let X be a compact set and let f(x) be a continuous function on X. Then $\min\{f(x):x\in X\}$ exists.

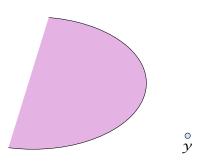
(without proof)

Lemma 35 (Projection Lemma)

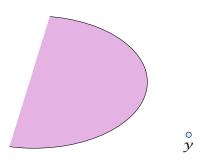
Let $X \subseteq \mathbb{R}^m$ be a non-empty convex set, and let $y \notin X$. Then there exist $x^* \in X$ with minimum distance from y. Moreover for all $x \in X$ we have $(y - x^*)^T (x - x^*) \le 0$.



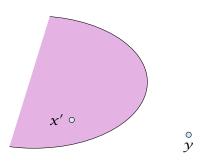
- ▶ Define f(x) = ||y x||.
- ▶ We want to apply Weierstrass but *X* may not be bounded.
- $\triangleright X \neq \emptyset$. Hence, there exists $x' \in X$.
- ▶ Define $X' = \{x \in X \mid \|y x\| \le \|y x'\|\}$. This set is closed and bounded.
- Applying Weierstrass gives the existence.



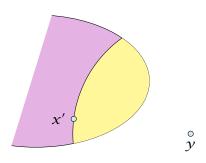
- ▶ Define f(x) = ||y x||.
- We want to apply Weierstrass but X may not be bounded.
- $X \neq \emptyset$. Hence, there exists $x' \in X$.
- ▶ Define $X' = \{x \in X \mid \|y x\| \le \|y x'\|\}$. This set is closed and bounded.
- Applying Weierstrass gives the existence.



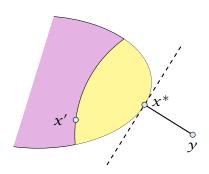
- ▶ Define f(x) = ||y x||.
- ightharpoonup We want to apply Weierstrass but X may not be bounded.
- $\blacktriangleright X \neq \emptyset$. Hence, there exists $\chi' \in X$.
- ▶ Define $X' = \{x \in X \mid \|y x\| \le \|y x'\|\}$. This set is closed and bounded.
- Applying Weierstrass gives the existence.



- ▶ Define f(x) = ||y x||.
- ightharpoonup We want to apply Weierstrass but X may not be bounded.
- ▶ $X \neq \emptyset$. Hence, there exists $x' \in X$.
- ▶ Define $X' = \{x \in X \mid \|y x\| \le \|y x'\|\}$. This set is closed and bounded.
- Applying Weierstrass gives the existence



- ▶ Define f(x) = ||y x||.
- ightharpoonup We want to apply Weierstrass but X may not be bounded.
- ▶ $X \neq \emptyset$. Hence, there exists $x' \in X$.
- ▶ Define $X' = \{x \in X \mid \|y x\| \le \|y x'\|\}$. This set is closed and bounded.
- Applying Weierstrass gives the existence.



 x^* is minimum. Hence $||y - x^*||^2 \le ||y - x||^2$ for all $x \in X$.

 x^* is minimum. Hence $||y - x^*||^2 \le ||y - x||^2$ for all $x \in X$.

 x^* is minimum. Hence $||y - x^*||^2 \le ||y - x||^2$ for all $x \in X$.

$$\|y - x^*\|^2$$

 x^* is minimum. Hence $||y - x^*||^2 \le ||y - x||^2$ for all $x \in X$.

$$\|y - x^*\|^2 \le \|y - x^* - \epsilon(x - x^*)\|^2$$

 x^* is minimum. Hence $||y - x^*||^2 \le ||y - x||^2$ for all $x \in X$.

$$||y - x^*||^2 \le ||y - x^* - \epsilon(x - x^*)||^2$$

$$= ||y - x^*||^2 + \epsilon^2 ||x - x^*||^2 - 2\epsilon(y - x^*)^T (x - x^*)$$

 x^* is minimum. Hence $||y - x^*||^2 \le ||y - x||^2$ for all $x \in X$.

By convexity: $x \in X$ then $x^* + \epsilon(x - x^*) \in X$ for all $0 \le \epsilon \le 1$.

$$||y - x^*||^2 \le ||y - x^* - \epsilon(x - x^*)||^2$$

$$= ||y - x^*||^2 + \epsilon^2 ||x - x^*||^2 - 2\epsilon(y - x^*)^T (x - x^*)$$

Hence, $(y - x^*)^T (x - x^*) \le \frac{1}{2} \epsilon ||x - x^*||^2$.

 x^* is minimum. Hence $||y - x^*||^2 \le ||y - x||^2$ for all $x \in X$.

By convexity: $x \in X$ then $x^* + \epsilon(x - x^*) \in X$ for all $0 \le \epsilon \le 1$.

$$||y - x^*||^2 \le ||y - x^* - \epsilon(x - x^*)||^2$$

$$= ||y - x^*||^2 + \epsilon^2 ||x - x^*||^2 - 2\epsilon(y - x^*)^T (x - x^*)$$

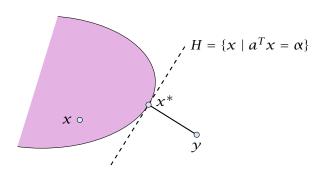
Hence,
$$(y - x^*)^T (x - x^*) \le \frac{1}{2} \epsilon ||x - x^*||^2$$
.

Letting $\epsilon \to 0$ gives the result.

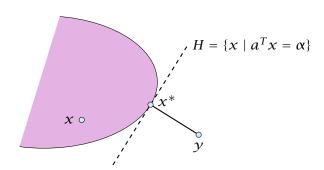
Theorem 36 (Separating Hyperplane)

Let $X \subseteq \mathbb{R}^m$ be a non-empty closed convex set, and let $y \notin X$. Then there exists a separating hyperplane $\{x \in \mathbb{R} : a^Tx = \alpha\}$ where $a \in \mathbb{R}^m$, $\alpha \in \mathbb{R}$ that separates y from X. $(a^Ty < \alpha; a^Tx \ge \alpha)$ for all $x \in X$

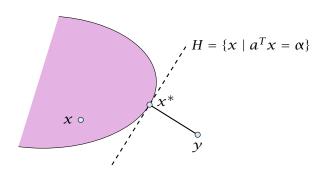
- Let $x^* \in X$ be closest point to y in X.
- ▶ By previous lemma $(y x^*)^T (x x^*) \le 0$ for all $x \in X$.
- Choose $a = (x^* y)$ and $\alpha = a^T x^*$.
- For $x \in X$: $a^T(x x^*) \ge 0$, and, hence, $a^Tx \ge \alpha$.
- ► Also, $a^T y = a^T (x^* a) = \alpha ||a||^2 < \alpha$



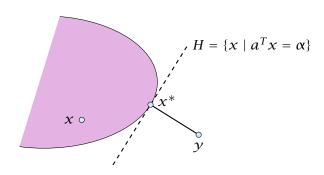
- Let $x^* \in X$ be closest point to y in X.
- ▶ By previous lemma $(y x^*)^T (x x^*) \le 0$ for all $x \in X$.
- Choose $a = (x^* y)$ and $\alpha = a^T x^*$.
- For $x \in X$: $a^T(x x^*) \ge 0$, and, hence, $a^Tx \ge \alpha$.
- Also, $a^T y = a^T (x^* a) = \alpha ||a||^2 < \alpha$



- Let $x^* \in X$ be closest point to y in X.
- ▶ By previous lemma $(y x^*)^T (x x^*) \le 0$ for all $x \in X$.
- Choose $a = (x^* y)$ and $\alpha = a^T x^*$.
- For $x \in X$: $a^T(x x^*) \ge 0$, and, hence, $a^Tx \ge \alpha$.
- ► Also, $a^T y = a^T (x^* a) = \alpha ||a||^2 < \alpha$

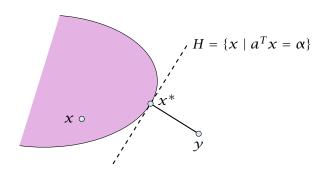


- Let $x^* \in X$ be closest point to y in X.
- ▶ By previous lemma $(y x^*)^T (x x^*) \le 0$ for all $x \in X$.
- Choose $a = (x^* y)$ and $\alpha = a^T x^*$.
- For $x \in X$: $a^T(x x^*) \ge 0$, and, hence, $a^Tx \ge \alpha$.
- ► Also, $a^T y = a^T (x^* a) = \alpha ||a||^2 < \alpha$



Proof of the Hyperplane Lemma

- Let $x^* \in X$ be closest point to y in X.
- ▶ By previous lemma $(y x^*)^T (x x^*) \le 0$ for all $x \in X$.
- Choose $a = (x^* y)$ and $\alpha = a^T x^*$.
- For $x \in X$: $a^T(x x^*) \ge 0$, and, hence, $a^Tx \ge \alpha$.
- Also, $a^T y = a^T (x^* a) = \alpha ||a||^2 < \alpha$



Lemma 37 (Farkas Lemma)

Let A be an $m \times n$ matrix, $b \in \mathbb{R}^m$. Then exactly one of the following statements holds.

- **1.** $\exists x \in \mathbb{R}^n$ with Ax = b, $x \ge 0$
- **2.** $\exists y \in \mathbb{R}^m$ with $A^T y \ge 0$, $b^T y < 0$

Assume \hat{x} satisfies 1. and \hat{y} satisfies 2. Then

$$0 > y^Tb = y^TAx \ge 0$$

Hence, at most one of the statements can hold.

Lemma 37 (Farkas Lemma)

Let A be an $m \times n$ matrix, $b \in \mathbb{R}^m$. Then exactly one of the following statements holds.

- **1.** $\exists x \in \mathbb{R}^n$ with Ax = b, $x \ge 0$
- **2.** $\exists y \in \mathbb{R}^m$ with $A^T y \ge 0$, $b^T y < 0$

Assume \hat{x} satisfies 1. and \hat{y} satisfies 2. Then

$$0 > y^T b = y^T A x \ge 0$$

Hence, at most one of the statements can hold

Lemma 37 (Farkas Lemma)

Let A be an $m \times n$ matrix, $b \in \mathbb{R}^m$. Then exactly one of the following statements holds.

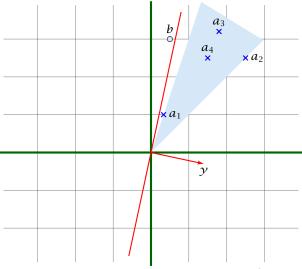
- 1. $\exists x \in \mathbb{R}^n$ with Ax = b, $x \ge 0$
- **2.** $\exists y \in \mathbb{R}^m$ with $A^T y \ge 0$, $b^T y < 0$

Assume \hat{x} satisfies 1. and \hat{y} satisfies 2. Then

$$0 > y^T b = y^T A x \ge 0$$

Hence, at most one of the statements can hold.

Farkas Lemma



If b is not in the cone generated by the columns of A, there exists a hyperplane y that separates b from the cone.

Now, assume that 1. does not hold

Consider $S = \{Ax : x \ge 0\}$ so that S closed, convex, $b \notin S$.

We want to show that there is y with $A^Ty \ge 0$, $b^Ty < 0$.

Let y be a hyperplane that separates b from S. Hence, $y^Tb < \alpha$ and $y^Ts \ge \alpha$ for all $s \in S$.

$$0 \in S \Rightarrow \alpha \le 0 \Rightarrow y^T b < 0$$

Now, assume that 1. does not hold.

Consider $S = \{Ax : x \ge 0\}$ so that S closed, convex, $b \notin S$.

We want to show that there is y with $A^Ty \ge 0$, $b^Ty < 0$.

Let y be a hyperplane that separates b from S. Hence, $y^Tb < \alpha$ and $y^Ts \ge \alpha$ for all $s \in S$.

$$0 \in S \Rightarrow \alpha \le 0 \Rightarrow y^T b < 0$$

Now, assume that 1. does not hold.

Consider $S = \{Ax : x \ge 0\}$ so that S closed, convex, $b \notin S$.

We want to show that there is y with $A^Ty \ge 0$, $b^Ty < 0$.

Let y be a hyperplane that separates b from S. Hence, $y^Tb < \alpha$ and $y^Ts \ge \alpha$ for all $s \in S$.

$$0 \in S \Rightarrow \alpha \le 0 \Rightarrow y^T b < 0$$

Now, assume that 1. does not hold.

Consider $S = \{Ax : x \ge 0\}$ so that S closed, convex, $b \notin S$.

We want to show that there is y with $A^Ty \ge 0$, $b^Ty < 0$.

Let y be a hyperplane that separates b from S. Hence, $y^Tb < \alpha$ and $y^Ts \ge \alpha$ for all $s \in S$.

 $0 \in S \Rightarrow \alpha \le 0 \Rightarrow y^T b < 0$

Now, assume that 1. does not hold.

Consider $S = \{Ax : x \ge 0\}$ so that S closed, convex, $b \notin S$.

We want to show that there is y with $A^Ty \ge 0$, $b^Ty < 0$.

Let y be a hyperplane that separates b from S. Hence, $y^Tb < \alpha$ and $y^Ts \ge \alpha$ for all $s \in S$.

 $0 \in S \Rightarrow \alpha \le 0 \Rightarrow y^T b < 0$

Now, assume that 1. does not hold.

Consider $S = \{Ax : x \ge 0\}$ so that S closed, convex, $b \notin S$.

We want to show that there is y with $A^Ty \ge 0$, $b^Ty < 0$.

Let y be a hyperplane that separates b from S. Hence, $y^Tb < \alpha$ and $y^Ts \ge \alpha$ for all $s \in S$.

$$0 \in S \Rightarrow \alpha \le 0 \Rightarrow y^T b < 0$$

Now, assume that 1. does not hold.

Consider $S = \{Ax : x \ge 0\}$ so that S closed, convex, $b \notin S$.

We want to show that there is y with $A^Ty \ge 0$, $b^Ty < 0$.

Let y be a hyperplane that separates b from S. Hence, $y^Tb < \alpha$ and $y^Ts \ge \alpha$ for all $s \in S$.

$$0 \in S \Rightarrow \alpha \le 0 \Rightarrow y^T b < 0$$

Now, assume that 1. does not hold.

Consider $S = \{Ax : x \ge 0\}$ so that S closed, convex, $b \notin S$.

We want to show that there is y with $A^Ty \ge 0$, $b^Ty < 0$.

Let y be a hyperplane that separates b from S. Hence, $y^Tb < \alpha$ and $y^Ts \ge \alpha$ for all $s \in S$.

$$0 \in S \Rightarrow \alpha \le 0 \Rightarrow y^T b < 0$$

Lemma 38 (Farkas Lemma; different version)

Let A be an $m \times n$ matrix, $b \in \mathbb{R}^m$. Then exactly one of the following statements holds.

- **1.** $\exists x \in \mathbb{R}^n$ with $Ax \leq b$, $x \geq 0$
- **2.** $\exists y \in \mathbb{R}^m$ with $A^T y \ge 0$, $b^T y < 0$, $y \ge 0$

Rewrite the conditions

1.
$$\exists x \in \mathbb{R}^n \text{ with } \begin{bmatrix} A \ I \end{bmatrix} \cdot \begin{bmatrix} x \\ s \end{bmatrix} = b, \ x \ge 0, \ s \ge 0$$

2.
$$\exists y \in \mathbb{R}^m \text{ with } \begin{bmatrix} A^T \\ I \end{bmatrix} y \ge 0, b^T y < 0$$

Lemma 38 (Farkas Lemma; different version)

Let A be an $m \times n$ matrix, $b \in \mathbb{R}^m$. Then exactly one of the following statements holds.

- **1.** $\exists x \in \mathbb{R}^n$ with $Ax \leq b$, $x \geq 0$
- **2.** $\exists y \in \mathbb{R}^m$ with $A^T y \ge 0$, $b^T y < 0$, $y \ge 0$

Rewrite the conditions:

1.
$$\exists x \in \mathbb{R}^n \text{ with } [A \ I] \cdot \begin{vmatrix} x \\ s \end{vmatrix} = b, \ x \ge 0, \ s \ge 0$$

2.
$$\exists y \in \mathbb{R}^m \text{ with } \begin{bmatrix} A^T \\ I \end{bmatrix} y \ge 0, b^T y < 0$$

$$P: z = \max\{c^T x \mid Ax \le b, x \ge 0\}$$

D:
$$w = \min\{b^T y \mid A^T y \ge c, y \ge 0\}$$

Theorem 39 (Strong Duality)

Let P and D be a primal dual pair of linear programs, and let z and w denote the optimal solution to P and D, respectively (i.e., P and D are non-empty). Then

$$z = w$$
.

 $z \le w$: follows from weak duality

 $z \le w$: follows from weak duality

 $z \geq w$:

 $z \le w$: follows from weak duality

 $z \ge w$:

We show $z < \alpha$ implies $w < \alpha$.

 $z \le w$: follows from weak duality

 $z \geq w$:

We show $z < \alpha$ implies $w < \alpha$.

$$\exists x \in \mathbb{R}^n$$
s.t.
$$Ax \leq b$$

$$-c^T x \leq -\alpha$$

$$x \geq 0$$

 $z \leq w$: follows from weak duality

 $z \geq w$:

We show $z < \alpha$ implies $w < \alpha$.

$$\exists x \in \mathbb{R}^n$$
s.t.
$$Ax \leq b$$

$$-c^T x \leq -\alpha$$

$$x \geq 0$$

$$\exists y \in \mathbb{R}^m; v \in \mathbb{R}$$
s.t. $A^T y - cv \ge 0$

$$b^T y - \alpha v < 0$$

$$y, v \ge 0$$

 $z \leq w$: follows from weak duality

 $z \geq w$:

We show $z < \alpha$ implies $w < \alpha$.

$$\exists x \in \mathbb{R}^n$$
s.t.
$$Ax \leq b$$

$$-c^T x \leq -\alpha$$

$$x \geq 0$$

$$\exists y \in \mathbb{R}^m; v \in \mathbb{R}$$
s.t. $A^T y - cv \ge 0$

$$b^T y - \alpha v < 0$$

$$y, v \ge 0$$

From the definition of α we know that the first system is infeasible; hence the second must be feasible.

$$\exists y \in \mathbb{R}^m; v \in \mathbb{R}$$
s.t. $A^T y - cv \ge 0$

$$b^T y - \alpha v < 0$$

$$y, v \ge 0$$

$$\exists y \in \mathbb{R}^m; v \in \mathbb{R}$$
s.t. $A^T y - cv \ge 0$

$$b^T y - \alpha v < 0$$

$$y, v \ge 0$$

If the solution y, v has v = 0 we have that

$$\exists y \in \mathbb{R}^m$$
s.t. $A^T y \ge 0$

$$b^T y < 0$$

$$y \ge 0$$

is feasible.

$$\exists y \in \mathbb{R}^{m}; v \in \mathbb{R}$$
s.t. $A^{T}y - cv \geq 0$

$$b^{T}y - \alpha v < 0$$

$$y, v \geq 0$$

If the solution y, v has v = 0 we have that

$$\exists y \in \mathbb{R}^m$$
s.t. $A^T y \ge 0$

$$b^T y < 0$$

$$y \ge 0$$

is feasible. By Farkas lemma this gives that LP ${\it P}$ is infeasible. Contradiction to the assumption of the lemma.

Hence, there exists a solution y, v with v > 0.

We can rescale this solution (scaling both y and v) s.t. v = 1.

Then y is feasible for the dual but $b^Ty<lpha.$ This means that w<lpha.

Hence, there exists a solution y, v with v > 0.

We can rescale this solution (scaling both y and v) s.t. v = 1.

Then y is feasible for the dual but $b^Ty < \alpha$. This means that $w < \alpha$.

Hence, there exists a solution y, v with v > 0.

We can rescale this solution (scaling both y and v) s.t. v=1.

Then y is feasible for the dual but $b^Ty < \alpha$. This means that $w < \alpha$.

Hence, there exists a solution y, v with v > 0.

We can rescale this solution (scaling both y and v) s.t. v = 1.

Then y is feasible for the dual but $b^Ty < \alpha$. This means that $w < \alpha$.

Definition 40 (Linear Programming Problem (LP))

Let $A \in \mathbb{Q}^{m \times n}$, $b \in \mathbb{Q}^m$, $c \in \mathbb{Q}^n$, $\alpha \in \mathbb{Q}$. Does there exist $x \in \mathbb{Q}^n$ s.t. Ax = b, $x \ge 0$, $c^T x \ge \alpha$?

Questions:

- ► Is LP in NP?
- Is LP in co-NP? yes!
- ► Is LP in P?

Proof

Definition 40 (Linear Programming Problem (LP))

Let $A \in \mathbb{Q}^{m \times n}$, $b \in \mathbb{Q}^m$, $c \in \mathbb{Q}^n$, $\alpha \in \mathbb{Q}$. Does there exist $x \in \mathbb{Q}^n$ s.t. Ax = b, $x \ge 0$, $c^T x \ge \alpha$?

Questions:

- ► Is LP in NP?
- Is LP in co-NP? yes!
- ► Is LP in P?

Proof:

- Given a primal maximization problem P and a parameter α . Suppose that $\alpha > \operatorname{opt}(P)$.
- We can prove this by providing an optimal basis for the dual.
- A verifier can check that the associated dual solution fulfills all dual constraints and that it has dual cost $< \alpha$.

Definition 40 (Linear Programming Problem (LP))

Let $A \in \mathbb{Q}^{m \times n}$, $b \in \mathbb{Q}^m$, $c \in \mathbb{Q}^n$, $\alpha \in \mathbb{Q}$. Does there exist $x \in \mathbb{Q}^n$ s.t. Ax = b, $x \ge 0$, $c^Tx \ge \alpha$?

Questions:

- ► Is LP in NP?
- Is LP in co-NP? yes!
- Is LP in P?

Proof:

- Given a primal maximization problem P and a parameter α . Suppose that $\alpha > \operatorname{opt}(P)$.
- We can prove this by providing an optimal basis for the dual.
- A verifier can check that the associated dual solution fulfills all dual constraints and that it has dual cost $< \alpha$.



Definition 40 (Linear Programming Problem (LP))

Let $A \in \mathbb{Q}^{m \times n}$, $b \in \mathbb{Q}^m$, $c \in \mathbb{Q}^n$, $\alpha \in \mathbb{Q}$. Does there exist $x \in \mathbb{Q}^n$ s.t. Ax = b, $x \ge 0$, $c^T x \ge \alpha$?

Questions:

- ► Is LP in NP?
- Is LP in co-NP? yes!
- ► Is LP in P?

Proof:

- Given a primal maximization problem P and a parameter α . Suppose that $\alpha > \operatorname{opt}(P)$.
- We can prove this by providing an optimal basis for the dual.
- A verifier can check that the associated dual solution fulfills all dual constraints and that it has dual cost $< \alpha$.



Complementary Slackness

Lemma 41

Assume a linear program $P = \max\{c^Tx \mid Ax \leq b; x \geq 0\}$ has solution x^* and its dual $D = \min\{b^Ty \mid A^Ty \geq c; y \geq 0\}$ has solution y^* .

- **1.** If $x_j^* > 0$ then the *j*-th constraint in *D* is tight.
- **2.** If the *j*-th constraint in *D* is not tight than $x_i^* = 0$.
- **3.** If $y_i^* > 0$ then the *i*-th constraint in *P* is tight.
- **4.** If the *i*-th constraint in *P* is not tight than $y_i^* = 0$.

Complementary Slackness

Lemma 41

Assume a linear program $P = \max\{c^Tx \mid Ax \leq b; x \geq 0\}$ has solution x^* and its dual $D = \min\{b^Ty \mid A^Ty \geq c; y \geq 0\}$ has solution y^* .

- **1.** If $x_j^* > 0$ then the *j*-th constraint in *D* is tight.
- **2.** If the *j*-th constraint in *D* is not tight than $x_i^* = 0$.
- **3.** If $y_i^* > 0$ then the *i*-th constraint in *P* is tight.
- **4.** If the *i*-th constraint in *P* is not tight than $y_i^* = 0$.

If we say that a variable x_j^* (y_i^*) has slack if $x_j^* > 0$ ($y_i^* > 0$), (i.e., the corresponding variable restriction is not tight) and a contraint has slack if it is not tight, then the above says that for a primal-dual solution pair it is not possible that a constraint **and** its corresponding (dual) variable has slack.

Proof: Complementary Slackness

Analogous to the proof of weak duality we obtain

$$c^Tx^* \leq y^{*T}Ax^* \leq b^Ty^*$$

Proof: Complementary Slackness

Analogous to the proof of weak duality we obtain

$$c^T x^* \le y^{*T} A x^* \le b^T y^*$$

Because of strong duality we then get

$$c^T x^* = y^{*T} A x^* = b^T y^*$$

This gives e.g.

$$\sum_{i} (y^T A - c^T)_j x_j^* = 0$$

Proof: Complementary Slackness

Analogous to the proof of weak duality we obtain

$$c^T x^* \le y^{*T} A x^* \le b^T y^*$$

Because of strong duality we then get

$$c^T x^* = y^{*T} A x^* = b^T y^*$$

This gives e.g.

$$\sum_{j} (y^T A - c^T)_j x_j^* = 0$$

From the constraint of the dual it follows that $y^TA \ge c^T$. Hence the left hand side is a sum over the product of non-negative numbers. Hence, if e.g. $(y^TA - c^T)_j > 0$ (the j-th constraint in the dual is not tight) then $x_j = 0$ (2.). The result for (1./3./4.) follows similarly.

Brewer: find mix of ale and beer that maximizes profits

max
$$13a + 23b$$

s.t. $5a + 15b \le 480$
 $4a + 4b \le 160$
 $35a + 20b \le 1190$
 $a, b \ge 0$

Entrepeneur: buy resources from brewer at minimum cost C, H, M: unit price for corn, hops and malt.

min
$$480C$$
 + $160H$ + $1190M$
s.t. $5C$ + $4H$ + $35M \ge 13$
 $15C$ + $4H$ + $20M \ge 23$
 $C, H, M \ge 0$

Note that brewer won't sell (at least not all) if e.g. 5C + 4H + 35M < 13 as then brewing ale would be advantageous.

Brewer: find mix of ale and beer that maximizes profits

max
$$13a + 23b$$

s.t. $5a + 15b \le 480$
 $4a + 4b \le 160$
 $35a + 20b \le 1190$
 $a, b \ge 0$

► Entrepeneur: buy resources from brewer at minimum cost *C*, *H*, *M*: unit price for corn, hops and malt.

min
$$480C$$
 + $160H$ + $1190M$
s.t. $5C$ + $4H$ + $35M \ge 13$
 $15C$ + $4H$ + $20M \ge 23$
 $C, H, M \ge 0$

Note that brewer won't sell (at least not all) if e.g. 5C + 4H + 35M < 13 as then brewing ale would be advantageous.

Brewer: find mix of ale and beer that maximizes profits

max
$$13a + 23b$$

s.t. $5a + 15b \le 480$
 $4a + 4b \le 160$
 $35a + 20b \le 1190$
 $a, b \ge 0$

► Entrepeneur: buy resources from brewer at minimum cost *C*, *H*, *M*: unit price for corn, hops and malt.

min
$$480C$$
 + $160H$ + $1190M$
s.t. $5C$ + $4H$ + $35M \ge 13$
 $15C$ + $4H$ + $20M \ge 23$
 $C, H, M \ge 0$

Note that brewer won't sell (at least not all) if e.g. 5C+4H+35M<13 as then brewing ale would be advantageous.

Marginal Price:

- How much money is the brewer willing to pay for additional amount of Corn, Hops, or Malt?
- ▶ We are interested in the marginal price, i.e., what happens if we increase the amount of Corn, Hops, and Malt by ε_C , ε_H , and ε_M , respectively.

The profit increases to $\max\{c^Tx \mid Ax \leq b + \varepsilon; x \geq 0\}$. Because of strong duality this is equal to

min
$$(b^T + \epsilon^T)y$$

s.t. $A^Ty \ge c$
 $y \ge 0$

Marginal Price:

- How much money is the brewer willing to pay for additional amount of Corn, Hops, or Malt?
- ▶ We are interested in the marginal price, i.e., what happens if we increase the amount of Corn, Hops, and Malt by ε_C , ε_H , and ε_M , respectively.

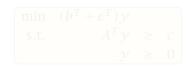
The profit increases to $\max\{c^Tx\mid Ax\leq b+\epsilon; x\geq 0\}$. Because of strong duality this is equal to



Marginal Price:

- How much money is the brewer willing to pay for additional amount of Corn, Hops, or Malt?
- ▶ We are interested in the marginal price, i.e., what happens if we increase the amount of Corn, Hops, and Malt by ε_C , ε_H , and ε_M , respectively.

The profit increases to $\max\{c^Tx\mid Ax\leq b+\varepsilon; x\geq 0\}$. Because of strong duality this is equal to



Marginal Price:

- How much money is the brewer willing to pay for additional amount of Corn, Hops, or Malt?
- ▶ We are interested in the marginal price, i.e., what happens if we increase the amount of Corn, Hops, and Malt by ε_C , ε_H , and ε_M , respectively.

The profit increases to $\max\{c^Tx\mid Ax\leq b+\varepsilon; x\geq 0\}$. Because of strong duality this is equal to

min
$$(b^T + \epsilon^T)y$$

s.t. $A^T y \ge c$
 $y \ge 0$

If ϵ is "small" enough then the optimum dual solution y^* might not change. Therefore the profit increases by $\sum_i arepsilon_i y_i^*$.

Therefore we can interpret the dual variables as marginal prices.

- If the brewer has slack of some resource (e.g. com) then he is not willing to pay anything for it (corresponding dual
 - variable is zero).
- If the dual variable for some resource is non-zero, then an increase of this resource increases the profit of the brewer. Hence it makes no sense to have left-aways of this resource are
 - Therefore its slack must be zero

If ϵ is "small" enough then the optimum dual solution y^* might not change. Therefore the profit increases by $\sum_i \epsilon_i y_i^*$.

Therefore we can interpret the dual variables as marginal prices.

- If the brewer has stack of some resource (e.g. com) then hee
- is not willing to pay anything for it (corresponding dual
- If the dual variable for some resource is non-zero, then any
- increase of this resource increases the profit of the brewer.
- Hence, it makes no sense to have left-overs of this resource...
- Therefore its slack must be zero

If ϵ is "small" enough then the optimum dual solution y^* might not change. Therefore the profit increases by $\sum_i \epsilon_i y_i^*$.

Therefore we can interpret the dual variables as marginal prices.

If ϵ is "small" enough then the optimum dual solution y^* might not change. Therefore the profit increases by $\sum_i \epsilon_i y_i^*$.

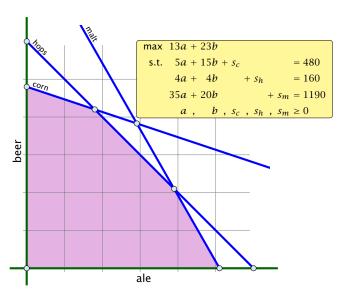
Therefore we can interpret the dual variables as marginal prices.

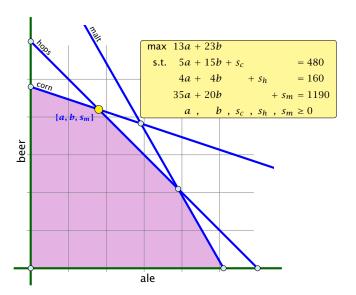
- If the brewer has slack of some resource (e.g. corn) then he is not willing to pay anything for it (corresponding dual variable is zero).
- ▶ If the dual variable for some resource is non-zero, then an increase of this resource increases the profit of the brewer. Hence, it makes no sense to have left-overs of this resource Therefore its slack must be zero.

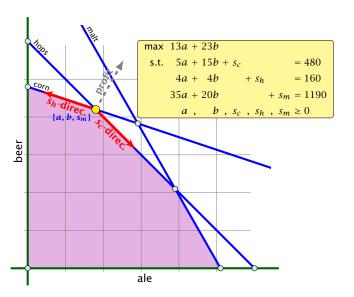
If ϵ is "small" enough then the optimum dual solution y^* might not change. Therefore the profit increases by $\sum_i \epsilon_i y_i^*$.

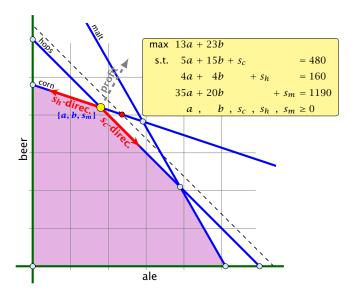
Therefore we can interpret the dual variables as marginal prices.

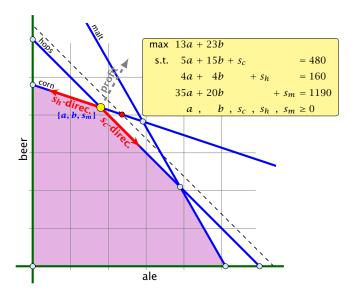
- ▶ If the brewer has slack of some resource (e.g. corn) then he is not willing to pay anything for it (corresponding dual variable is zero).
- ▶ If the dual variable for some resource is non-zero, then an increase of this resource increases the profit of the brewer. Hence, it makes no sense to have left-overs of this resource. Therefore its slack must be zero.

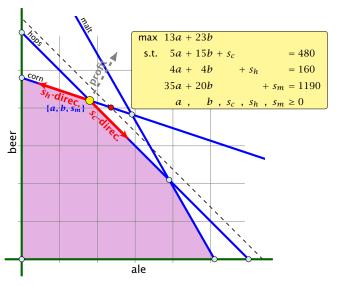




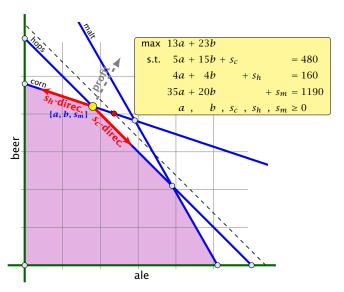








The change in profit when increasing hops by one unit is $= c_B^T A_B^{-1} e_h$.



The change in profit when increasing hops by one unit is

$$=\underbrace{c_B^T A_B^{-1}}_{v^*} e_h.$$

Of course, the previous argument about the increase in the primal objective only holds for the non-degenerate case.

If the optimum basis is degenerate then increasing the supply of one resource may not allow the objective value to increase.

Definition 42

An (s,t)-flow in a (complete) directed graph $G=(V,V\times V,c)$ is a function $f:V\times V\mapsto \mathbb{R}^+_0$ that satisfies

1. For each edge (x, y)

$$0 \le f_{xy} \le c_{xy} \ .$$

(capacity constraints)

2. For each $v \in V \setminus \{s, t\}$

$$\sum_{x} f_{vx} = \sum_{x} f_{xv}$$

(flow conservation constraints)



Definition 42

An (s,t)-flow in a (complete) directed graph $G=(V,V\times V,c)$ is a function $f:V\times V\mapsto \mathbb{R}^+_0$ that satisfies

1. For each edge (x, y)

$$0 \le f_{xy} \le c_{xy} .$$

(capacity constraints)

2. For each $v \in V \setminus \{s, t\}$

$$\sum_{x} f_{vx} = \sum_{x} f_{xv} .$$

(flow conservation constraints)



Definition 43

The value of an (s, t)-flow f is defined as

$$\operatorname{val}(f) = \sum_{x} f_{sx} - \sum_{x} f_{xs}$$
.

Maximum Flow Problem:

Find an (s,t)-flow with maximum value

Definition 43

The value of an (s, t)-flow f is defined as

$$val(f) = \sum_{x} f_{sx} - \sum_{x} f_{xs} .$$

Maximum Flow Problem:

Find an (s, t)-flow with maximum value.

max
$$\sum_{z} f_{sz} - \sum_{z} f_{zs}$$
s.t. $\forall (z, w) \in V \times V$
$$f_{zw} \leq c_{zw} \quad \ell_{zw}$$

$$\forall w \neq s, t \quad \sum_{z} f_{zw} - \sum_{z} f_{wz} = 0 \qquad p_{w}$$

$$f_{zw} \geq 0$$

max
$$\sum_{z} f_{sz} - \sum_{z} f_{zs}$$
s.t. $\forall (z, w) \in V \times V$
$$f_{zw} \leq c_{zw} \quad \ell_{zw}$$

$$\forall w \neq s, t \quad \sum_{z} f_{zw} - \sum_{z} f_{wz} = 0 \quad p_{w}$$

$$f_{zw} \geq 0$$

min		$\sum_{(xy)} c_{xy} \ell_{xy}$		
s.t.	$f_{xy}(x, y \neq s, t)$:	$1\ell_{xy}-1p_x+1p_y$	≥	0
	$f_{sy} (y \neq s, t)$:	$1\ell_{sy}$ $+1p_y$	≥	1
	$f_{xs}(x \neq s,t)$:	$1\ell_{xs}-1p_x$	≥	-1
	$f_{ty} (y \neq s, t)$:	$1\ell_{ty}$ $+1p_y$	≥	0
	$f_{xt} (x \neq s, t)$:	$1\ell_{xt}$ – $1p_x$	≥	0
	f_{st} :	$1\ell_{st}$	≥	1
	f_{ts} :	$1\ell_{ts}$	≥	-1
		ℓ_{xy}	≥	0

min		$\sum_{(xy)} c_{xy} \ell_{xy}$	
s.t.	$f_{xy}(x, y \neq s, t)$:	$1\ell_{xy}-1p_x+1p_y \geq$	0
	$f_{sy} (y \neq s, t)$:	$1\ell_{sy}$ - $1+1p_y \ge$	0
	$f_{xs}(x \neq s, t)$:	$1\ell_{xs}-1p_x+1 \geq$	0
	$f_{ty} (y \neq s, t)$:	$1\ell_{ty}$ $ 0+1p_y \ge$	0
	f_{xt} $(x \neq s, t)$:	$1\ell_{xt}-1p_x+0 \ge$	0
	f_{st} :	$1\ell_{st}$ - $1+$ $0 \ge$	0
	f_{ts} :	$1\ell_{ts}$ $ 0+$ $1 \geq$	0
		$\ell_{xy} \geq$	0

$$\begin{array}{llll} & & \sum_{(xy)} c_{xy} \ell_{xy} \\ \text{s.t.} & f_{xy} \; (x,y \neq s,t) \colon & 1\ell_{xy} - 1p_x + 1p_y \; \geq \; 0 \\ & f_{sy} \; (y \neq s,t) \colon & 1\ell_{sy} - \; p_s + 1p_y \; \geq \; 0 \\ & f_{xs} \; (x \neq s,t) \colon & 1\ell_{xs} - 1p_x + \; p_s \; \geq \; 0 \\ & f_{ty} \; (y \neq s,t) \colon & 1\ell_{ty} - \; p_t + 1p_y \; \geq \; 0 \\ & f_{xt} \; (x \neq s,t) \colon & 1\ell_{xt} - 1p_x + \; p_t \; \geq \; 0 \\ & f_{st} \colon & 1\ell_{st} - \; p_s + \; p_t \; \geq \; 0 \\ & f_{ts} \colon & 1\ell_{ts} - \; p_t + \; p_s \; \geq \; 0 \\ & \ell_{xy} \; \geq \; 0 \end{array}$$

with $p_t = 0$ and $p_s = 1$.

min
$$\sum_{(xy)} c_{xy} \ell_{xy}$$
s.t. f_{xy} : $1\ell_{xy} - 1p_x + 1p_y \ge 0$

$$\ell_{xy} \ge 0$$

$$p_s = 1$$

$$p_t = 0$$

We can interpret the ℓ_{xy} value as assigning a length to every edge.

The value p_x for a variable, then can be seen as the distance of x to t (where the distance from s to t is required to be 1 since $p_s = 1$).

The constraint $p_X \le \ell_{XY} + p_Y$ then simply follows from triangle inequality $(d(x,t) \le d(x,y) + d(y,t) \Rightarrow d(x,t) \le \ell_{XY} + d(y,t))$

min
$$\sum_{(xy)} c_{xy} \ell_{xy}$$
s.t. f_{xy} : $1\ell_{xy} - 1p_x + 1p_y \ge 0$

$$\ell_{xy} \ge 0$$

$$p_s = 1$$

$$p_t = 0$$

We can interpret the ℓ_{xy} value as assigning a length to every edge.

The value p_x for a variable, then can be seen as the distance of x to t (where the distance from s to t is required to be 1 since $p_s = 1$).

The constraint $p_X \le \ell_{XY} + p_Y$ then simply follows from triangle inequality $(d(x,t) \le d(x,y) + d(y,t) \Rightarrow d(x,t) \le \ell_{XY} + d(y,t))$

min
$$\sum_{(xy)} c_{xy} \ell_{xy}$$
s.t. f_{xy} : $1\ell_{xy} - 1p_x + 1p_y \ge 0$

$$\ell_{xy} \ge 0$$

$$p_s = 1$$

$$p_t = 0$$

We can interpret the ℓ_{xy} value as assigning a length to every edge.

The value p_x for a variable, then can be seen as the distance of x to t (where the distance from s to t is required to be 1 since $p_s = 1$).

The constraint $p_X \le \ell_{XY} + p_Y$ then simply follows from triangle inequality $(d(x,t) \le d(x,y) + d(y,t) \Rightarrow d(x,t) \le \ell_{XY} + d(y,t))$

min
$$\sum_{(xy)} c_{xy} \ell_{xy}$$
s.t. f_{xy} : $1\ell_{xy} - 1p_x + 1p_y \ge 0$

$$\ell_{xy} \ge 0$$

$$p_s = 1$$

$$p_t = 0$$

We can interpret the ℓ_{xy} value as assigning a length to every edge.

The value p_x for a variable, then can be seen as the distance of x to t (where the distance from s to t is required to be 1 since $p_s = 1$).

The constraint $p_x \le \ell_{xy} + p_y$ then simply follows from triangle inequality $(d(x,t) \le d(x,y) + d(y,t) \Rightarrow d(x,t) \le \ell_{xy} + d(y,t))$.

One can show that there is an optimum LP-solution for the dual problem that gives an integral assignment of variables.

This means $p_X = 1$ or $p_X = 0$ for our case. This gives rise to a cut in the graph with vertices having value 1 on one side and the other vertices on the other side. The objective function then evaluates the capacity of this cut.

This shows that the Maxflow/Mincut theorem follows from linear programming duality.

One can show that there is an optimum LP-solution for the dual problem that gives an integral assignment of variables.

This means $p_{\chi}=1$ or $p_{\chi}=0$ for our case. This gives rise to a cut in the graph with vertices having value 1 on one side and the other vertices on the other side. The objective function then evaluates the capacity of this cut.

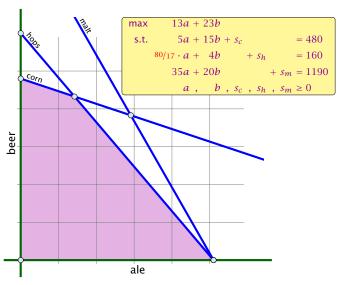
This shows that the Maxflow/Mincut theorem follows from linear programming duality.

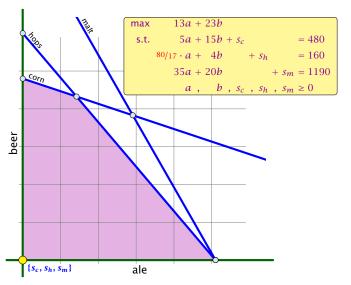
One can show that there is an optimum LP-solution for the dual problem that gives an integral assignment of variables.

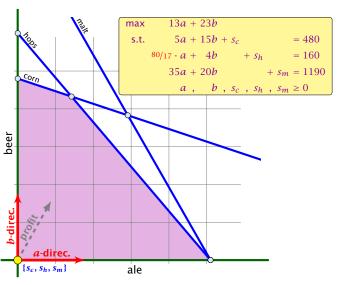
This means $p_{\chi}=1$ or $p_{\chi}=0$ for our case. This gives rise to a cut in the graph with vertices having value 1 on one side and the other vertices on the other side. The objective function then evaluates the capacity of this cut.

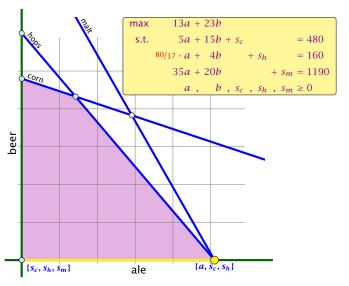
This shows that the Maxflow/Mincut theorem follows from linear programming duality.

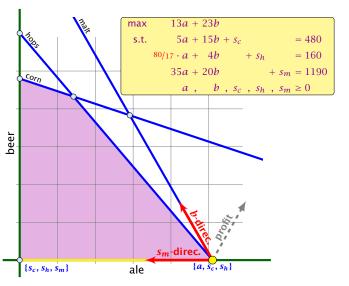
If a basis variable is 0 in the basic feasible solution then we may not make progress during an iteration of simplex.

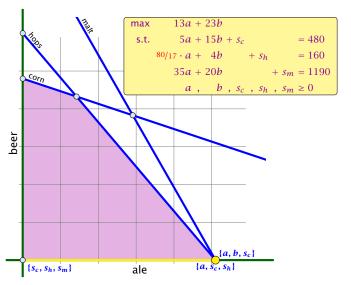


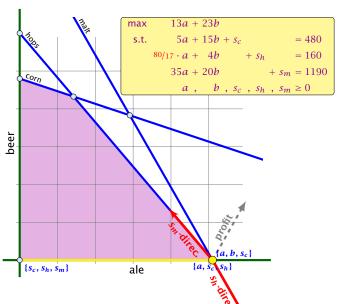


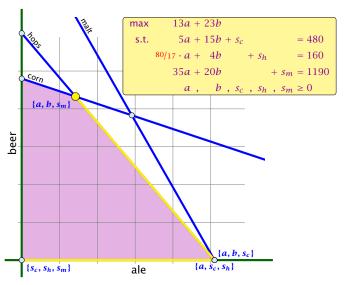


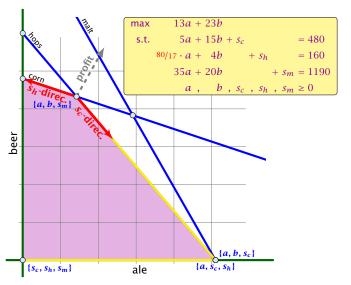












If a basis variable is 0 in the basic feasible solution then we may not make progress during an iteration of simplex.

Idea:

Given feasible LP := $\max\{c^Tx, Ax = b; x \ge 0\}$. Change it into LP' := $\max\{c^Tx, Ax = b', x \ge 0\}$ such that

```
III' is feasible
```

```
If a set it of basis variables corresponds to an (i.e. a in a) then it corresponds to an infeasible basis in
```

(note that columns if a vare linearly moependent)



If a basis variable is 0 in the basic feasible solution then we may not make progress during an iteration of simplex.

Idea:

Given feasible LP := $\max\{c^Tx, Ax = b; x \ge 0\}$. Change it into LP' := $\max\{c^Tx, Ax = b', x \ge 0\}$ such that

If a basis variable is 0 in the basic feasible solution then we may not make progress during an iteration of simplex.

Idea:

Given feasible LP := $\max\{c^Tx, Ax = b; x \ge 0\}$. Change it into LP' := $\max\{c^Tx, Ax = b', x \ge 0\}$ such that

LP' is feasible

II. If a set B of basis variables corresponds to an infeasible basis (i.e. $A_B^{-1}b \not \geq 0$) then B corresponds to an infeasible basis in LP' (note that columns in A_B are linearly independent).

III. LP' has no degenerate basic solutions

If a basis variable is 0 in the basic feasible solution then we may not make progress during an iteration of simplex.

Idea:

Given feasible LP := $\max\{c^Tx, Ax = b; x \ge 0\}$. Change it into LP' := $\max\{c^Tx, Ax = b', x \ge 0\}$ such that

- LP' is feasible
- II. If a set B of basis variables corresponds to an infeasible basis (i.e. $A_B^{-1}b \not\geq 0$) then B corresponds to an infeasible basis in LP' (note that columns in A_B are linearly independent).
- III. LP' has no degenerate basic solutions

If a basis variable is 0 in the basic feasible solution then we may not make progress during an iteration of simplex.

Idea:

Given feasible LP := $\max\{c^Tx, Ax = b; x \ge 0\}$. Change it into LP' := $\max\{c^Tx, Ax = b', x \ge 0\}$ such that

- LP' is feasible
- II. If a set B of basis variables corresponds to an infeasible basis (i.e. $A_B^{-1}b \not \geq 0$) then B corresponds to an infeasible basis in LP' (note that columns in A_B are linearly independent).
- III. LP' has no degenerate basic solutions

Perturbation

Let B be index set of some basis with basic solution

$$x_B^* = A_B^{-1}b \ge 0, x_N^* = 0$$
 (i.e. *B* is feasible)

Fix

$$b':=b+A_Begin{pmatrix}arepsilon\ arepsilon\ arepsilon m\end{pmatrix}$$
 for $arepsilon>0$. .

This is the perturbation that we are using.

Perturbation

Let B be index set of some basis with basic solution

$$x_B^* = A_B^{-1}b \ge 0, x_N^* = 0$$
 (i.e. *B* is feasible)

Fix

$$b':=b+A_B\left(egin{array}{c} arepsilon \ arepsilon \ arepsilon \end{array}
ight) ext{ for } arepsilon>0 \ .$$

This is the perturbation that we are using.

The new LP is feasible because the set B of basis variables provides a feasible basis:

$$A_B^{-1} \left(b + A_B \begin{pmatrix} \varepsilon \\ \vdots \\ \varepsilon^m \end{pmatrix} \right) = \chi_B^* + \begin{pmatrix} \varepsilon \\ \vdots \\ \varepsilon^m \end{pmatrix} \ge 0$$

The new LP is feasible because the set B of basis variables provides a feasible basis:

$$A_B^{-1} \left(b + A_B \begin{pmatrix} \varepsilon \\ \vdots \\ \varepsilon^m \end{pmatrix} \right) = x_B^* + \begin{pmatrix} \varepsilon \\ \vdots \\ \varepsilon^m \end{pmatrix} \ge 0 .$$

Let \tilde{B} be a non-feasible basis. This means $(A_{\tilde{B}}^{-1}b)_i<0$ for some row i.

Let \tilde{B} be a non-feasible basis. This means $(A_{\tilde{B}}^{-1}b)_i<0$ for some row i.

Then for small enough $\epsilon > 0$

$$\left(A_{\tilde{B}}^{-1}\left(b+A_{B}\begin{pmatrix}\varepsilon\\\vdots\\\varepsilon^{m}\end{pmatrix}\right)\right)_{i}$$

Let \tilde{B} be a non-feasible basis. This means $(A_{\tilde{B}}^{-1}b)_i<0$ for some row i.

Then for small enough $\epsilon > 0$

$$\left(A_{\tilde{B}}^{-1}\left(b+A_{B}\begin{pmatrix}\varepsilon\\\vdots\\\varepsilon^{m}\end{pmatrix}\right)\right)_{i} = (A_{\tilde{B}}^{-1}b)_{i} + \left(A_{\tilde{B}}^{-1}A_{B}\begin{pmatrix}\varepsilon\\\vdots\\\varepsilon^{m}\end{pmatrix}\right)_{i} < 0$$

Let \tilde{B} be a non-feasible basis. This means $(A_{\tilde{B}}^{-1}b)_i<0$ for some row i.

Then for small enough $\epsilon > 0$

$$\left(A_{\tilde{B}}^{-1}\left(b+A_{B}\begin{pmatrix}\varepsilon\\\vdots\\\varepsilon^{m}\end{pmatrix}\right)\right)_{i} = (A_{\tilde{B}}^{-1}b)_{i} + \left(A_{\tilde{B}}^{-1}A_{B}\begin{pmatrix}\varepsilon\\\vdots\\\varepsilon^{m}\end{pmatrix}\right)_{i} < 0$$

Hence, \tilde{B} is not feasible.

Let \tilde{B} be a basis. It has an associated solution

$$\chi_{\tilde{B}}^* = A_{\tilde{B}}^{-1}b + A_{\tilde{B}}^{-1}A_B \begin{pmatrix} \varepsilon \\ \vdots \\ \varepsilon^m \end{pmatrix}$$

in the perturbed instance.

We can view each component of the vector as a polynom with variable arepsilon of degree at most m.

 $A_{ ilde{B}}^{-1}A_B$ has rank m. Therefore no polynom is 0.

A polynom of degree at most m has at most m roots (Nullstellen).

Let \tilde{B} be a basis. It has an associated solution

$$\chi_{\tilde{B}}^* = A_{\tilde{B}}^{-1}b + A_{\tilde{B}}^{-1}A_B \begin{pmatrix} \varepsilon \\ \vdots \\ \varepsilon^m \end{pmatrix}$$

in the perturbed instance.

We can view each component of the vector as a polynom with variable ε of degree at most m.

 $A_{ ilde{B}}^{-1}A_B$ has rank m. Therefore no polynom is 0.

A polynom of degree at most m has at most m roots (Nullstellen).

Let \tilde{B} be a basis. It has an associated solution

$$\chi_{\tilde{B}}^* = A_{\tilde{B}}^{-1}b + A_{\tilde{B}}^{-1}A_B \begin{pmatrix} \varepsilon \\ \vdots \\ \varepsilon^m \end{pmatrix}$$

in the perturbed instance.

We can view each component of the vector as a polynom with variable ε of degree at most m.

 $A_{\tilde{B}}^{-1}A_B$ has rank m. Therefore no polynom is 0.

A polynom of degree at most m has at most m roots (Nullstellen).

Let \tilde{B} be a basis. It has an associated solution

$$\chi_{\tilde{B}}^* = A_{\tilde{B}}^{-1}b + A_{\tilde{B}}^{-1}A_B \begin{pmatrix} \varepsilon \\ \vdots \\ \varepsilon^m \end{pmatrix}$$

in the perturbed instance.

We can view each component of the vector as a polynom with variable ε of degree at most m.

 $A_{\tilde{B}}^{-1}A_B$ has rank m. Therefore no polynom is 0.

A polynom of degree at most m has at most m roots (Nullstellen).

Let \tilde{B} be a basis. It has an associated solution

$$\chi_{\tilde{B}}^* = A_{\tilde{B}}^{-1}b + A_{\tilde{B}}^{-1}A_B \begin{pmatrix} \varepsilon \\ \vdots \\ \varepsilon^m \end{pmatrix}$$

in the perturbed instance.

We can view each component of the vector as a polynom with variable ε of degree at most m.

 $A_{\tilde{B}}^{-1}A_B$ has rank m. Therefore no polynom is 0.

A polynom of degree at most m has at most m roots (Nullstellen).

Let \tilde{B} be a basis. It has an associated solution

$$\chi_{\tilde{B}}^* = A_{\tilde{B}}^{-1}b + A_{\tilde{B}}^{-1}A_B \begin{pmatrix} \varepsilon \\ \vdots \\ \varepsilon^m \end{pmatrix}$$

in the perturbed instance.

We can view each component of the vector as a polynom with variable ε of degree at most m.

 $A_{\tilde{B}}^{-1}A_B$ has rank m. Therefore no polynom is 0.

A polynom of degree at most m has at most m roots (Nullstellen).

Since, there are no degeneracies Simplex will terminate when run on LP^\prime .

Since, there are no degeneracies Simplex will terminate when run on LP' .

If it terminates because the reduced cost vector fulfills

$$\tilde{c} = (c^T - c_B^T A_B^{-1} A) \leq 0$$

then we have found an optimal basis.

Since, there are no degeneracies Simplex will terminate when run on LP' .

If it terminates because the reduced cost vector fulfills

$$\tilde{c} = (c^T - c_B^T A_B^{-1} A) \leq 0$$

then we have found an optimal basis. Note that this basis is also optimal for LP, as the above constraint does not depend on b.

Since, there are no degeneracies Simplex will terminate when run on LP' .

If it terminates because the reduced cost vector fulfills

$$\tilde{c} = (c^T - c_B^T A_B^{-1} A) \leq 0$$

then we have found an optimal basis. Note that this basis is also optimal for LP, as the above constraint does not depend on b.

If it terminates because it finds a variable x_j with $\tilde{c}_j > 0$ for which the j-th basis direction d, fulfills $d \ge 0$ we know that LP' is unbounded. The basis direction does not depend on b. Hence, we also know that LP is unbounded.

Doing calculations with perturbed instances may be costly. Also the right choice of ε is difficult.

Idea

Simulate behaviour of LP' without explicitly doing a perturbation.

Doing calculations with perturbed instances may be costly. Also the right choice of ε is difficult.

Idea:

Simulate behaviour of LP' without explicitly doing a perturbation.

Doing calculations with perturbed instances may be costly. Also the right choice of ε is difficult.

Idea:

Simulate behaviour of LP' without explicitly doing a perturbation.

We choose the entering variable arbitrarily as before ($\tilde{c}_e > 0$, of course).

If we do not have a choice for the leaving variable then LP' and LP do the same (i.e., choose the same variable).

We choose the entering variable arbitrarily as before ($\tilde{c}_e > 0$, of course).

If we do not have a choice for the leaving variable then LP' and LP do the same (i.e., choose the same variable).

We choose the entering variable arbitrarily as before ($\tilde{c}_e > 0$, of course).

If we do not have a choice for the leaving variable then LP' and LP do the same (i.e., choose the same variable).

We choose the entering variable arbitrarily as before ($\tilde{c}_e > 0$, of course).

If we do not have a choice for the leaving variable then LP' and LP do the same (i.e., choose the same variable).

In the following we assume that $b \ge 0$. This can be obtained by replacing the initial system $(A \mid b)$ by $(A_B^{-1}A \mid A_B^{-1}b)$ where B is the index set of a feasible basis (found e.g. by the first phase of the Two-phase algorithm).

Then the perturbed instance is

$$b' = b + \begin{pmatrix} \varepsilon \\ \vdots \\ \varepsilon^m \end{pmatrix}$$

In the following we assume that $b \ge 0$. This can be obtained by replacing the initial system $(A \mid b)$ by $(A_B^{-1}A \mid A_B^{-1}b)$ where B is the index set of a feasible basis (found e.g. by the first phase of the Two-phase algorithm).

Then the perturbed instance is

$$b' = b + \begin{pmatrix} \varepsilon \\ \vdots \\ \varepsilon^m \end{pmatrix}$$

Matrix View

Let our linear program be

$$c_B^T x_B + c_N^T x_N = Z$$

$$A_B x_B + A_N x_N = b$$

$$x_B , x_N \ge 0$$

The simplex tableaux for basis B is

$$(c_N^T - c_B^T A_B^{-1} A_N) x_N = Z - c_B^T A_B^{-1} b$$

 $Ix_B + A_B^{-1} A_N x_N = A_B^{-1} b$
 $x_B , x_N \ge 0$

The BFS is given by $x_N = 0, x_B = A_B^{-1}b$.

If $(c_N^T - c_B^T A_B^{-1} A_N) \le 0$ we know that we have an optimum solution.



LP chooses an arbitrary leaving variable that has $\hat{A}_{\ell e}>0$ and minimizes

$$\theta_{\ell} = \frac{\hat{b}_{\ell}}{\hat{A}_{\ell e}} = \frac{(A_B^{-1}b)_{\ell}}{(A_B^{-1}A_{*e})_{\ell}}$$

 ℓ is the index of a leaving variable within B. This means if e.g. $B = \{1, 3, 7, 14\}$ and leaving variable is 3 then $\ell = 2$.

LP chooses an arbitrary leaving variable that has $\hat{A}_{\ell e}>0$ and minimizes

$$\theta_{\ell} = \frac{\hat{b}_{\ell}}{\hat{A}_{\ell e}} = \frac{(A_B^{-1}b)_{\ell}}{(A_B^{-1}A_{ke})_{\ell}}$$

 ℓ is the index of a leaving variable within B. This means if e.g. $B=\{1,3,7,14\}$ and leaving variable is 3 then $\ell=2$.

LP chooses an arbitrary leaving variable that has $\hat{A}_{\ell e}>0$ and minimizes

$$\theta_{\ell} = \frac{\hat{b}_{\ell}}{\hat{A}_{\ell e}} = \frac{(A_B^{-1}b)_{\ell}}{(A_B^{-1}A_{*e})_{\ell}} \ .$$

 ℓ is the index of a leaving variable within B. This means if e.g. $B = \{1, 3, 7, 14\}$ and leaving variable is 3 then $\ell = 2$.

LP chooses an arbitrary leaving variable that has $\hat{A}_{\ell e}>0$ and minimizes

$$\theta_{\ell} = \frac{\hat{b}_{\ell}}{\hat{A}_{\ell e}} = \frac{(A_B^{-1}b)_{\ell}}{(A_B^{-1}A_{*e})_{\ell}} \ .$$

 ℓ is the index of a leaving variable within B. This means if e.g. $B=\{1,3,7,14\}$ and leaving variable is 3 then $\ell=2$.

Definition 44

 $u \leq_{\mathsf{lex}} v$ if and only if the first component in which u and v differ fulfills $u_i \leq v_i$.

LP' chooses an index that minimizes

 θ_ℓ

LP' chooses an index that minimizes

$$\theta_{\ell} = \frac{\left(A_{B}^{-1} \left(b + \begin{pmatrix} \varepsilon \\ \vdots \\ \varepsilon^{m} \end{pmatrix}\right)\right)_{\ell}}{(A_{B}^{-1} A_{*e})_{\ell}}$$

LP' chooses an index that minimizes

$$\theta_{\ell} = \frac{\left(A_B^{-1} \left(b + \begin{pmatrix} \varepsilon \\ \vdots \\ \varepsilon^m \end{pmatrix}\right)\right)_{\ell}}{(A_B^{-1} A_{*e})_{\ell}} = \frac{\left(A_B^{-1} (b \mid I) \begin{pmatrix} 1 \\ \varepsilon \\ \vdots \\ \varepsilon^m \end{pmatrix}\right)_{\ell}}{(A_B^{-1} A_{*e})_{\ell}}$$

LP' chooses an index that minimizes

$$\theta_{\ell} = \frac{\left(A_{B}^{-1} \begin{pmatrix} b + \begin{pmatrix} \varepsilon \\ \vdots \\ \varepsilon^{m} \end{pmatrix} \right) \right)_{\ell}}{(A_{B}^{-1} A_{*e})_{\ell}} = \frac{\left(A_{B}^{-1} (b \mid I) \begin{pmatrix} 1 \\ \varepsilon \\ \vdots \\ \varepsilon^{m} \end{pmatrix} \right)_{\ell}}{(A_{B}^{-1} A_{*e})_{\ell}}$$

$$= \frac{\ell \text{-th row of } A_{B}^{-1} (b \mid I)}{(A_{B}^{-1} A_{*e})_{\ell}} \begin{pmatrix} 1 \\ \varepsilon \\ \vdots \\ \varepsilon^{m} \end{pmatrix}$$

This means you can choose the variable/row ℓ for which the vector

$$\frac{\ell\text{-th row of }A_B^{-1}(b\mid I)}{(A_B^{-1}A_{*e})_{\ell}}$$

is lexicographically minimal.

Of course only including rows with $(A_B^{-1}A_{*e})_{\ell} > 0$.

This technique guarantees that your pivoting is the same as in the perturbed case. This guarantees that cycling does not occur.

This means you can choose the variable/row ℓ for which the vector

$$\frac{\ell\text{-th row of }A_B^{-1}(b\mid I)}{(A_B^{-1}A_{*e})_{\ell}}$$

is lexicographically minimal.

Of course only including rows with $(A_B^{-1}A_{*e})_{\ell} > 0$.

This technique guarantees that your pivoting is the same as in the perturbed case. This guarantees that cycling does not occur.

This means you can choose the variable/row ℓ for which the vector

$$\frac{\ell\text{-th row of }A_B^{-1}(b\mid I)}{(A_B^{-1}A_{*e})_{\ell}}$$

is lexicographically minimal.

Of course only including rows with $(A_B^{-1}A_{*e})_{\ell} > 0$.

This technique guarantees that your pivoting is the same as in the perturbed case. This guarantees that cycling does not occur.

Each iteration of Simplex can be implemented in polynomial time.

Each iteration of Simplex can be implemented in polynomial time.

If we use lexicographic pivoting we know that Simplex requires at most $\binom{n}{m}$ iterations, because it will not visit a basis twice.

Each iteration of Simplex can be implemented in polynomial time.

If we use lexicographic pivoting we know that Simplex requires at most $\binom{n}{m}$ iterations, because it will not visit a basis twice.

The input size is $L \cdot n \cdot m$, where n is the number of variables, m is the number of constraints, and L is the length of the binary representation of the largest coefficient in the matrix A.

Each iteration of Simplex can be implemented in polynomial time.

If we use lexicographic pivoting we know that Simplex requires at most $\binom{n}{m}$ iterations, because it will not visit a basis twice.

The input size is $L \cdot n \cdot m$, where n is the number of variables, m is the number of constraints, and L is the length of the binary representation of the largest coefficient in the matrix A.

If we really require $\binom{n}{m}$ iterations then Simplex is not a polynomial time algorithm.

Each iteration of Simplex can be implemented in polynomial time.

If we use lexicographic pivoting we know that Simplex requires at most $\binom{n}{m}$ iterations, because it will not visit a basis twice.

The input size is $L \cdot n \cdot m$, where n is the number of variables, m is the number of constraints, and L is the length of the binary representation of the largest coefficient in the matrix A.

If we really require $\binom{n}{m}$ iterations then Simplex is not a polynomial time algorithm.

Can we obtain a better analysis?



Observation

Simplex visits every feasible basis at most once.

Observation

Simplex visits every feasible basis at most once.

However, also the number of feasible bases can be very large.

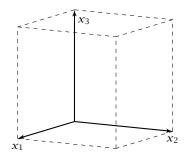
Example

$$\max c^{T}x$$
s.t. $0 \le x_{1} \le 1$

$$0 \le x_{2} \le 1$$

$$\vdots$$

$$0 \le x_{n} \le 1$$



2n constraint on n variables define an n-dimensional hypercube as feasible region.

The feasible region has 2^n vertices.

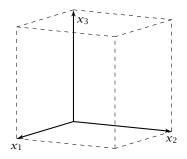
Example

$$\max c^{T}x$$
s.t. $0 \le x_{1} \le 1$

$$0 \le x_{2} \le 1$$

$$\vdots$$

$$0 \le x_{n} \le 1$$



However, Simplex may still run quickly as it usually does not visit all feasible bases.

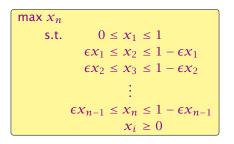
In the following we give an example of a feasible region for which there is a bad <u>Pivoting Rule</u>.

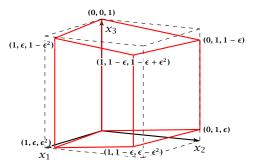
Pivoting Rule

A Pivoting Rule defines how to choose the entering and leaving variable for an iteration of Simplex.

In the non-degenerate case after choosing the entering variable the leaving variable is unique.

Klee Minty Cube





Observations

- We have 2n constraints, and 3n variables (after adding slack variables to every constraint).
- Every basis is defined by 2n variables, and n non-basic variables.
- There exist degenerate vertices.
- The degeneracies come from the non-negativity constraints, which are superfluous.
- In the following all variables x_i stay in the basis at all times
- ► Then, we can uniquely specify a basis by choosing for each variable whether it should be equal to its lower bound, or equal to its upper bound (the slack variable corresponding to the non-tight constraint is part of the basis).
- We can also simply identify each basis/vertex with the corresponding hypercube vertex obtained by letting $\epsilon \to 0$.

Observations

- We have 2n constraints, and 3n variables (after adding slack variables to every constraint).
- Every basis is defined by 2n variables, and n non-basic variables.
- There exist degenerate vertices.
- The degeneracies come from the non-negativity constraints, which are superfluous.
- In the following all variables x_i stay in the basis at all times
- ► Then, we can uniquely specify a basis by choosing for each variable whether it should be equal to its lower bound, or equal to its upper bound (the slack variable corresponding to the non-tight constraint is part of the basis).
- We can also simply identify each basis/vertex with the corresponding hypercube vertex obtained by letting $\epsilon \to 0$.

- We have 2n constraints, and 3n variables (after adding slack variables to every constraint).
- Every basis is defined by 2n variables, and n non-basic variables.
- There exist degenerate vertices.
- The degeneracies come from the non-negativity constraints, which are superfluous.
- In the following all variables x_i stay in the basis at all times
- Then, we can uniquely specify a basis by choosing for each variable whether it should be equal to its lower bound, or equal to its upper bound (the slack variable corresponding to the non-tight constraint is part of the basis).
- ▶ We can also simply identify each basis/vertex with the corresponding hypercube vertex obtained by letting $\epsilon \to 0$.

- We have 2n constraints, and 3n variables (after adding slack variables to every constraint).
- Every basis is defined by 2n variables, and n non-basic variables.
- There exist degenerate vertices.
- The degeneracies come from the non-negativity constraints, which are superfluous.
- In the following all variables x_i stay in the basis at all times
- ► Then, we can uniquely specify a basis by choosing for each variable whether it should be equal to its lower bound, or equal to its upper bound (the slack variable corresponding to the non-tight constraint is part of the basis).
- We can also simply identify each basis/vertex with the corresponding hypercube vertex obtained by letting $\epsilon \to 0$.

- We have 2n constraints, and 3n variables (after adding slack variables to every constraint).
- Every basis is defined by 2n variables, and n non-basic variables.
- There exist degenerate vertices.
- The degeneracies come from the non-negativity constraints, which are superfluous.
- In the following all variables x_i stay in the basis at all times.
- Then, we can uniquely specify a basis by choosing for each variable whether it should be equal to its lower bound, or equal to its upper bound (the slack variable corresponding to the non-tight constraint is part of the basis).
- We can also simply identify each basis/vertex with the corresponding hypercube vertex obtained by letting $\epsilon \to 0$.

- We have 2n constraints, and 3n variables (after adding slack variables to every constraint).
- Every basis is defined by 2n variables, and n non-basic variables.
- There exist degenerate vertices.
- The degeneracies come from the non-negativity constraints, which are superfluous.
- In the following all variables x_i stay in the basis at all times.
- ▶ Then, we can uniquely specify a basis by choosing for each variable whether it should be equal to its lower bound, or equal to its upper bound (the slack variable corresponding to the non-tight constraint is part of the basis).
- We can also simply identify each basis/vertex with the corresponding hypercube vertex obtained by letting $\epsilon \to 0$.

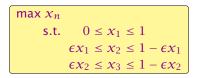
- We have 2n constraints, and 3n variables (after adding slack variables to every constraint).
- Every basis is defined by 2n variables, and n non-basic variables.
- There exist degenerate vertices.
- The degeneracies come from the non-negativity constraints, which are superfluous.
- In the following all variables x_i stay in the basis at all times.
- ► Then, we can uniquely specify a basis by choosing for each variable whether it should be equal to its lower bound, or equal to its upper bound (the slack variable corresponding to the non-tight constraint is part of the basis).
- ▶ We can also simply identify each basis/vertex with the corresponding hypercube vertex obtained by letting $\epsilon \to 0$.

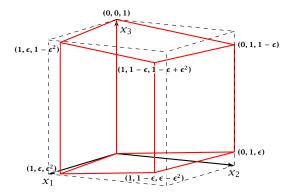
- In the following we specify a sequence of bases (identified by the corresponding hypercube node) along which the objective function strictly increases.
- ▶ The basis (0,...,0,1) is the unique optimal basis.
- Our sequence S_n starts at (0, ..., 0) ends with (0, ..., 0, 1) and visits every node of the hypercube.
- An unfortunate Pivoting Rule may choose this sequence, and, hence, require an exponential number of iterations.

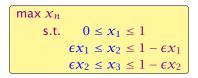
- In the following we specify a sequence of bases (identified by the corresponding hypercube node) along which the objective function strictly increases.
- ▶ The basis (0, ..., 0, 1) is the unique optimal basis.
- Our sequence S_n starts at (0, ..., 0) ends with (0, ..., 0, 1) and visits every node of the hypercube.
- An unfortunate Pivoting Rule may choose this sequence, and, hence, require an exponential number of iterations.

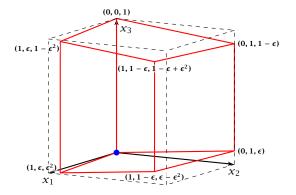
- In the following we specify a sequence of bases (identified by the corresponding hypercube node) along which the objective function strictly increases.
- ▶ The basis (0, ..., 0, 1) is the unique optimal basis.
- Our sequence S_n starts at (0, ..., 0) ends with (0, ..., 0, 1) and visits every node of the hypercube.
- An unfortunate Pivoting Rule may choose this sequence, and, hence, require an exponential number of iterations.

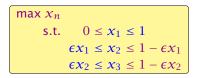
- In the following we specify a sequence of bases (identified by the corresponding hypercube node) along which the objective function strictly increases.
- ▶ The basis (0, ..., 0, 1) is the unique optimal basis.
- Our sequence S_n starts at (0, ..., 0) ends with (0, ..., 0, 1) and visits every node of the hypercube.
- An unfortunate Pivoting Rule may choose this sequence, and, hence, require an exponential number of iterations.

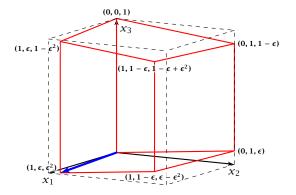


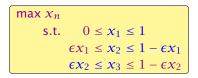


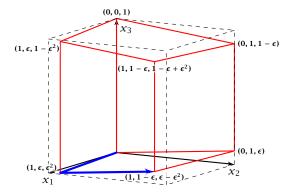


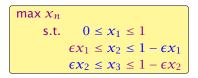


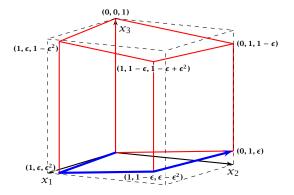


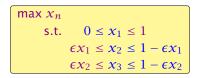


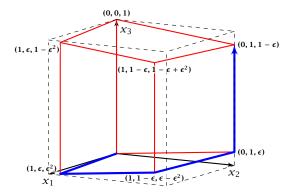


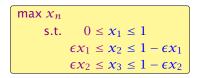


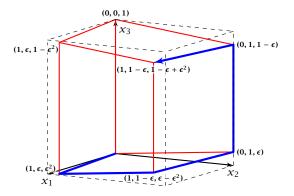


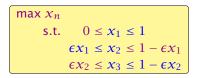


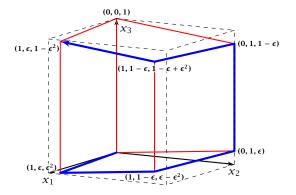


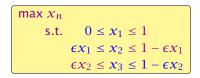


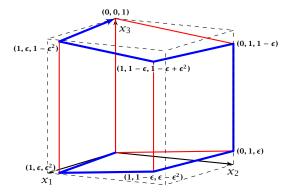












The sequence S_n that visits every node of the hypercube is defined recursively

$$(0,...,0,0,0)$$
 $\begin{cases} S_{n-1} \\ (0,...,0,1,0) \\ \\ (0,...,0,1,1) \\ \\ S_{n-1} \\ \\ (0,...,0,0,1) \end{cases}$

The non-recursive case is $S_1 = 0 \rightarrow 1$

Lemma 45

The objective value x_n is increasing along path S_n .

Proof by induction:

n=1: obvious, since $S_1=0 \rightarrow 1$, and 1>0.

 $n-1 \rightarrow n$

Lemma 45

The objective value x_n is increasing along path S_n .

Proof by induction:

```
n = 1: obvious, since S_1 = 0 \rightarrow 1, and 1 > 0.
```

 $n-1 \rightarrow n$

Lemma 45

The objective value x_n is increasing along path S_n .

Proof by induction:

n = 1: obvious, since $S_1 = 0 \rightarrow 1$, and 1 > 0.

 $n-1 \rightarrow n$

Lemma 45

The objective value x_n is increasing along path S_n .

$$n = 1$$
: obvious, since $S_1 = 0 \rightarrow 1$, and $1 > 0$.

$$n-1 \rightarrow n$$

- ▶ For the first part the value of $x_n = \epsilon x_{n-1}$.
- ▶ By induction hypothesis x_{n-1} is increasing along S_{n-1} , hence, also x_n .
- ▶ Going from (0,...,0,1,0) to (0,...,0,1,1) increases x_n for small enough ϵ .
- ▶ For the remaining path S_{n-1}^{rev} we have $x_n = 1 \epsilon x_{n-1}$.
- By induction hypothesis x_{n-1} is increasing along S_{n-1} , hence $-\epsilon x_{n-1}$ is increasing along S_{n-1}^{rev} .

Lemma 45

The objective value x_n is increasing along path S_n .

$$n = 1$$
: obvious, since $S_1 = 0 \rightarrow 1$, and $1 > 0$.

$$n-1 \rightarrow n$$

- ▶ For the first part the value of $x_n = \epsilon x_{n-1}$.
- By induction hypothesis x_{n-1} is increasing along S_{n-1} , hence, also x_n .
- ▶ Going from (0,...,0,1,0) to (0,...,0,1,1) increases x_n for small enough ϵ .
- ▶ For the remaining path S_{n-1}^{rev} we have $x_n = 1 \epsilon x_{n-1}$.
- By induction hypothesis x_{n-1} is increasing along S_{n-1} , hence $-\epsilon x_{n-1}$ is increasing along S_{n-1}^{rev} .

Lemma 45

The objective value x_n is increasing along path S_n .

```
n = 1: obvious, since S_1 = 0 \rightarrow 1, and 1 > 0.
```

$$n-1 \rightarrow n$$

- For the first part the value of $x_n = \epsilon x_{n-1}$.
- By induction hypothesis x_{n-1} is increasing along S_{n-1} , hence, also x_n .
- ▶ Going from (0,...,0,1,0) to (0,...,0,1,1) increases x_n for small enough ϵ .
- ▶ For the remaining path S_{n-1}^{rev} we have $x_n = 1 \epsilon x_{n-1}$.
- By induction hypothesis x_{n-1} is increasing along S_{n-1} , hence $-\epsilon x_{n-1}$ is increasing along S_{n-1}^{rev} .

Lemma 45

The objective value x_n is increasing along path S_n .

```
n = 1: obvious, since S_1 = 0 \rightarrow 1, and 1 > 0.
```

$$n-1 \rightarrow n$$

- For the first part the value of $x_n = \epsilon x_{n-1}$.
- By induction hypothesis x_{n-1} is increasing along S_{n-1} , hence, also x_n .
- ▶ Going from (0,...,0,1,0) to (0,...,0,1,1) increases x_n for small enough ϵ .
- ► For the remaining path S_{n-1}^{rev} we have $x_n = 1 \epsilon x_{n-1}$.
- By induction hypothesis x_{n-1} is increasing along S_{n-1} , hence $-\epsilon x_{n-1}$ is increasing along S_{n-1}^{rev} .

Lemma 45

The objective value x_n is increasing along path S_n .

- n = 1: obvious, since $S_1 = 0 \rightarrow 1$, and 1 > 0.
- $n-1 \rightarrow n$
 - ▶ For the first part the value of $x_n = \epsilon x_{n-1}$.
 - ▶ By induction hypothesis x_{n-1} is increasing along S_{n-1} , hence, also x_n .
 - ▶ Going from (0,...,0,1,0) to (0,...,0,1,1) increases x_n for small enough ϵ .
 - ▶ For the remaining path S_{n-1}^{rev} we have $x_n = 1 \epsilon x_{n-1}$.
 - ▶ By induction hypothesis x_{n-1} is increasing along S_{n-1} , hence $-\epsilon x_{n-1}$ is increasing along S_{n-1}^{rev} .

Observation

The simplex algorithm takes at most $\binom{n}{m}$ iterations. Each iteration can be implemented in time $\mathcal{O}(mn)$.

In practise it usually takes a linear number of iterations.

Theorem

For almost all known deterministic pivoting rules (rules for choosing entering and leaving variables) there exist lower bounds that require the algorithm to have exponential running time $(\Omega(2^{\Omega(n)}))$ (e.g. Klee Minty 1972).

Theorem

For some standard randomized pivoting rules there exist subexponential lower bounds ($\Omega(2^{\Omega(n^{\alpha})})$ for $\alpha>0$) (Friedmann, Hansen, Zwick 2011).

Conjecture (Hirsch 1957)

The edge-vertex graph of an m-facet polytope in d-dimensional Euclidean space has diameter no more than m-d.

The conjecture has been proven wrong in 2010.

But the question whether the diameter is perhaps of the form $\mathcal{O}(\operatorname{poly}(m,d))$ is open.

- Suppose we want to solve $\min\{c^Tx \mid Ax \geq b; x \geq 0\}$, where $x \in \mathbb{R}^d$ and we have m constraints.
- In the worst-case Simplex runs in time roughly $\mathcal{O}(m(m+d){m+d \choose m}) \approx (m+d)^m$. (slightly better bounds or the running time exist, but will not be discussed here).
- If d is much smaller than m one can do a lot better.
- In the following we develop an algorithm with running time $\mathcal{O}(d! \cdot m)$, i.e., linear in m.

- ▶ Suppose we want to solve $\min\{c^Tx \mid Ax \geq b; x \geq 0\}$, where $x \in \mathbb{R}^d$ and we have m constraints.
- In the worst-case Simplex runs in time roughly $\mathcal{O}(m(m+d)\binom{m+d}{m}) \approx (m+d)^m$. (slightly better bounds on the running time exist, but will not be discussed here).
- ▶ If *d* is much smaller than *m* one can do a lot better
- In the following we develop an algorithm with running time $\mathcal{O}(d! \cdot m)$, i.e., linear in m.

- ▶ Suppose we want to solve $\min\{c^Tx \mid Ax \geq b; x \geq 0\}$, where $x \in \mathbb{R}^d$ and we have m constraints.
- In the worst-case Simplex runs in time roughly $\mathcal{O}(m(m+d)\binom{m+d}{m}) \approx (m+d)^m$. (slightly better bounds on the running time exist, but will not be discussed here).
- If d is much smaller than m one can do a lot better.
- ▶ In the following we develop an algorithm with running time $\mathcal{O}(d! \cdot m)$, i.e., linear in m.

- ▶ Suppose we want to solve $\min\{c^Tx \mid Ax \geq b; x \geq 0\}$, where $x \in \mathbb{R}^d$ and we have m constraints.
- In the worst-case Simplex runs in time roughly $\mathcal{O}(m(m+d)\binom{m+d}{m}) \approx (m+d)^m$. (slightly better bounds on the running time exist, but will not be discussed here).
- ▶ If *d* is much smaller than *m* one can do a lot better.
- In the following we develop an algorithm with running time $\mathcal{O}(d! \cdot m)$, i.e., linear in m.

Setting:

We assume an LP of the form

$$\begin{array}{cccc}
\min & c^T x \\
\text{s.t.} & Ax & \geq & b \\
& & x & \geq & 0
\end{array}$$

We assume that the LP is bounded.

Ensuring Conditions

Given a standard minimization LP

$$\begin{array}{cccc}
\min & c^T x \\
\text{s.t.} & Ax & \geq & b \\
& & x & \geq & 0
\end{array}$$

how can we obtain an LP of the required form?

Compute a lower bound on $c^T x$ for any basic feasible solution.

Let s denote the smallest common multiple of all denominators of entries in A, b.

Multiply entries in A,b by s to obtain integral entries. This does not change the feasible region.

Add slack variables to A; denote the resulting matrix with $ar{A}.$

Let s denote the smallest common multiple of all denominators of entries in A, b.

Multiply entries in A, b by s to obtain integral entries. This does not change the feasible region.

Add slack variables to A; denote the resulting matrix with \overline{A} .

Let s denote the smallest common multiple of all denominators of entries in A, b.

Multiply entries in A, b by s to obtain integral entries. This does not change the feasible region.

Add slack variables to A; denote the resulting matrix with \bar{A} .

Let s denote the smallest common multiple of all denominators of entries in A, b.

Multiply entries in A, b by s to obtain integral entries. This does not change the feasible region.

Add slack variables to A; denote the resulting matrix with \bar{A} .

Theorem 46 (Cramers Rule)

Let M be a matrix with $\det(M) \neq 0$. Then the solution to the system Mx = b is given by

$$x_i = \frac{\det(M_j)}{\det(M)}$$
 ,

where M_i is the matrix obtained from M by replacing the i-th column by the vector b.

Define

$$X_i = \begin{pmatrix} | & | & | & | \\ e_1 \cdots e_{i-1} & \mathbf{x} & e_{i+1} \cdots e_n \\ | & | & | & | \end{pmatrix}$$

Note that expanding along the *i*-th column gives that $det(X_i) = x_i$.

Further, we have

Hence,

$$x_i = \det(X_i) = \frac{\det(M_i)}{\det(M)}$$

Define

$$X_i = \begin{pmatrix} | & | & | & | \\ e_1 \cdots e_{i-1} & \mathbf{x} & e_{i+1} \cdots e_n \\ | & | & | & | \end{pmatrix}$$

Note that expanding along the *i*-th column gives that $det(X_i) = x_i$.

Further, we have

$$MX_i = \begin{pmatrix} & & & & & & & & & & & & \\ Me_1 & \cdots & Me_{i-1} & Mx & Me_{i+1} & \cdots & Me_n \\ & & & & & & & & & & \end{pmatrix} = M$$

Hence.

$$x_i = \det(X_i) = \frac{\det(M_i)}{\det(M)}$$

Define

$$X_i = \begin{pmatrix} | & | & | & | \\ e_1 \cdots e_{i-1} & \mathbf{x} & e_{i+1} \cdots e_n \\ | & | & | & | \end{pmatrix}$$

Note that expanding along the *i*-th column gives that $det(X_i) = x_i$.

Further, we have

$$MX_i = \begin{pmatrix} | & | & | & | \\ Me_1 \cdot \cdot \cdot \cdot Me_{i-1} & Mx & Me_{i+1} \cdot \cdot \cdot \cdot Me_n \\ | & | & | & | \end{pmatrix} = M_i$$

Hence.

$$x_i = \det(X_i) = \frac{\det(M_i)}{\det(M)}$$

Define

$$X_i = \begin{pmatrix} | & | & | & | \\ e_1 \cdots e_{i-1} & \mathbf{x} & e_{i+1} \cdots e_n \\ | & | & | & | \end{pmatrix}$$

Note that expanding along the *i*-th column gives that $det(X_i) = x_i$.

Further, we have

$$MX_i = \begin{pmatrix} | & | & | & | \\ Me_1 \cdot \cdot \cdot \cdot Me_{i-1} & Mx & Me_{i+1} \cdot \cdot \cdot \cdot Me_n \\ | & | & | & | \end{pmatrix} = M_i$$

Hence,

$$x_i = \det(X_i) = \frac{\det(M_i)}{\det(M)}$$

Let Z be the maximum absolute entry occuring in \bar{A} , \bar{b} or c. Let C denote the matrix obtained from \bar{A}_B by replacing the j-th column with vector \bar{b} (for some j).

Observe that

 $|\det(C)|$

Let Z be the maximum absolute entry occuring in \bar{A} , \bar{b} or c. Let C denote the matrix obtained from \bar{A}_B by replacing the j-th column with vector \bar{b} (for some j).

Observe that

$$|\det(C)| = \left| \sum_{\pi \in S_m} \operatorname{sgn}(\pi) \prod_{1 \le i \le m} C_{i\pi(i)} \right|$$

Let Z be the maximum absolute entry occuring in \bar{A} , \bar{b} or c. Let C denote the matrix obtained from \bar{A}_B by replacing the j-th column with vector \bar{b} (for some j).

Observe that

$$|\det(C)| = \left| \sum_{\pi \in S_m} \operatorname{sgn}(\pi) \prod_{1 \le i \le m} C_{i\pi(i)} \right|$$

$$\le \sum_{\pi \in S_m} \prod_{1 \le i \le m} |C_{i\pi(i)}|$$

Let Z be the maximum absolute entry occuring in \bar{A} , \bar{b} or c. Let C denote the matrix obtained from \bar{A}_B by replacing the j-th column with vector \bar{b} (for some j).

Observe that

$$|\det(C)| = \left| \sum_{\pi \in S_m} \operatorname{sgn}(\pi) \prod_{1 \le i \le m} C_{i\pi(i)} \right|$$

$$\le \sum_{\pi \in S_m} \prod_{1 \le i \le m} |C_{i\pi(i)}|$$

$$\le m! \cdot Z^m.$$

Alternatively, Hadamards inequality gives

|det(*C*)|

Alternatively, Hadamards inequality gives

$$|\det(C)| \le \prod_{i=1}^{m} ||C_{*i}||$$

Alternatively, Hadamards inequality gives

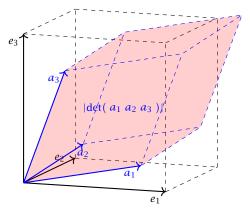
$$|\det(C)| \le \prod_{i=1}^{m} ||C_{*i}|| \le \prod_{i=1}^{m} (\sqrt{m}Z)$$

Alternatively, Hadamards inequality gives

$$|\det(C)| \le \prod_{i=1}^m ||C_{*i}|| \le \prod_{i=1}^m (\sqrt{m}Z)$$

$$\le m^{m/2}Z^m.$$

Hadamards Inequality



Hadamards inequality says that the volume of the red parallelepiped (Spat) is smaller than the volume in the black cube (if $||e_1|| = ||a_1||$, $||e_2|| = ||a_2||$, $||e_3|| = ||a_3||$).

Ensuring Conditions

Given a standard minimization LP

$$\begin{array}{cccc}
\min & c^T x \\
\text{s.t.} & Ax & \geq & b \\
& & x & \geq & 0
\end{array}$$

how can we obtain an LP of the required form?

► Compute a lower bound on c^Tx for any basic feasible solution. Add the constraint $c^Tx \ge -dZ(m! \cdot Z^m) - 1$. Note that this constraint is superfluous unless the LP is unbounded.

Ensuring Conditions

Compute an optimum basis for the new LP.

- ▶ If the cost is $c^T x = -(dZ)(m! \cdot Z^m) 1$ we know that the original LP is unbounded.
- Otw. we have an optimum basis.

In the following we use \mathcal{H} to denote the set of all constraints apart from the constraint $c^Tx \ge -dZ(m! \cdot Z^m) - 1$.

We give a routine SeidelLP(\mathcal{H},d) that is given a set \mathcal{H} of explicit, non-degenerate constraints over d variables, and minimizes c^Tx over all feasible points.

In addition it obeys the implicit constraint $c^T x \ge -(dZ)(m! \cdot Z^m) - 1$.

In the following we use \mathcal{H} to denote the set of all constraints apart from the constraint $c^Tx \geq -dZ(m! \cdot Z^m) - 1$.

We give a routine SeidelLP(\mathcal{H},d) that is given a set \mathcal{H} of explicit, non-degenerate constraints over d variables, and minimizes c^Tx over all feasible points.

In addition it obeys the implicit constraint $c^T x \ge -(dZ)(m! \cdot Z^m) - 1$.

In the following we use \mathcal{H} to denote the set of all constraints apart from the constraint $c^Tx \geq -dZ(m! \cdot Z^m) - 1$.

We give a routine SeidelLP(\mathcal{H},d) that is given a set \mathcal{H} of explicit, non-degenerate constraints over d variables, and minimizes c^Tx over all feasible points.

In addition it obeys the implicit constraint $c^T x \ge -(dZ)(m! \cdot Z^m) - 1$.

In the following we use \mathcal{H} to denote the set of all constraints apart from the constraint $c^Tx \geq -dZ(m! \cdot Z^m) - 1$.

We give a routine SeidelLP(\mathcal{H}, d) that is given a set \mathcal{H} of explicit, non-degenerate constraints over d variables, and minimizes $c^T x$ over all feasible points.

In addition it obeys the implicit constraint $c^Tx \ge -(dZ)(m! \cdot Z^m) - 1$.

Algorithm 1 SeidelLP(\mathcal{H}, d) 1: if d = 1 then solve 1-dimensional problem and return;

Algorithm 1 SeidelLP(\mathcal{H}, d) 1: **if** d = 1 **then** solve 1-dimensional problem and return;

- - 2: if $\mathcal{H} = \emptyset$ then return x on implicit constraint hyperplane

Algorithm 1 SeidelLP(\mathcal{H}, d) 1: **if** d = 1 **then** solve 1-dimensional problem and return;

- 2: **if** $\mathcal{H} = \emptyset$ **then** return x on implicit constraint hyperplane 3: choose random constraint $h \in \mathcal{H}$

- 1: **if** d = 1 **then** solve 1-dimensional problem and return;

- 2: **if** $\mathcal{H} = \emptyset$ **then** return x on implicit constraint hyperplane 3: choose random constraint $h \in \mathcal{H}$ 4: $\hat{\mathcal{H}} \leftarrow \mathcal{H} \setminus \{h\}$

- 1: **if** d = 1 **then** solve 1-dimensional problem and return;
- 2: **if** $\mathcal{H} = \emptyset$ **then** return x on implicit constraint hyperplane 3: choose random constraint $h \in \mathcal{H}$ 4: $\hat{\mathcal{H}} \leftarrow \mathcal{H} \setminus \{h\}$ 5: $\hat{x}^* \leftarrow \text{SeidelLP}(\hat{\mathcal{H}}, d)$

5: $\hat{x}^* \leftarrow \text{SeidelLP}(\hat{\mathcal{H}}, d)$

- 1: **if** d = 1 **then** solve 1-dimensional problem and return;
- 2: **if** $\mathcal{H} = \emptyset$ **then** return x on implicit constraint hyperplane
- 3: choose random constraint $h \in \mathcal{H}$

3: choose random constraint
$$h \in \mathcal{H}$$

4: $\hat{\mathcal{H}} \leftarrow \mathcal{H} \setminus \{h\}$

6: if \hat{x}^* = infeasible then return infeasible

- 2: if $\mathcal{H} = \emptyset$ then return x on implicit constraint hyperplane
- 3: choose random constraint $h \in \mathcal{H}$
- 4: $\hat{\mathcal{H}} \leftarrow \mathcal{H} \setminus \{h\}$
- 5: $\hat{x}^* \leftarrow \text{SeidelLP}(\hat{\mathcal{H}}, d)$

7: **if** \hat{x}^* fulfills h then return \hat{x}^*

- 6: if \hat{x}^* = infeasible then return infeasible

- 1: **if** d = 1 **then** solve 1-dimensional problem and return;

- 1: **if** d = 1 **then** solve 1-dimensional problem and return;
- 2: if $\mathcal{H} = \emptyset$ then return x on implicit constraint hyperplane

8: // optimal solution fulfills h with equality, i.e., $a_h^T x = b_h$

- 3: choose random constraint $h \in \mathcal{H}$
- 4: $\hat{\mathcal{H}} \leftarrow \mathcal{H} \setminus \{h\}$
- 5: $\hat{x}^* \leftarrow \text{SeidelLP}(\hat{\mathcal{H}}, d)$

- 6: if \hat{x}^* = infeasible then return infeasible
- 7: **if** \hat{x}^* fulfills h then return \hat{x}^*

- 1: **if** d = 1 **then** solve 1-dimensional problem and return;
- 2: if $\mathcal{H} = \emptyset$ then return x on implicit constraint hyperplane

3: choose random constraint
$$h \in \mathcal{H}$$

- 4: $\hat{\mathcal{H}} \leftarrow \mathcal{H} \setminus \{h\}$ 5: $\hat{x}^* \leftarrow \text{SeidelLP}(\hat{\mathcal{H}}, d)$

 - 6: if \hat{x}^* = infeasible then return infeasible
 - 7: **if** \hat{x}^* fulfills h then return \hat{x}^*
 - 8: // optimal solution fulfills h with equality, i.e., $a_h^T x = b_h$
 - 9: solve $a_h^T x = b_h$ for some variable x_ℓ ;
- 10: eliminate x_{ℓ} in constraints from $\hat{\mathcal{H}}$ and in implicit constr.;

- 1: if d = 1 then solve 1-dimensional problem and return;
- 2: if $\mathcal{H} = \emptyset$ then return x on implicit constraint hyperplane
- 3: choose random constraint $h \in \mathcal{H}$
 - 4: $\hat{\mathcal{H}} \leftarrow \mathcal{H} \setminus \{h\}$
 - 5: $\hat{x}^* \leftarrow \text{SeidelLP}(\hat{\mathcal{H}}, d)$
 - 6: if \hat{x}^* = infeasible then return infeasible
 - 7: **if** \hat{x}^* fulfills h then return \hat{x}^*

11: $\hat{x}^* \leftarrow \text{SeidelLP}(\hat{\mathcal{H}}, d-1)$

- 9: solve $a_h^T x = b_h$ for some variable x_ℓ ; 10: eliminate x_{ℓ} in constraints from $\hat{\mathcal{H}}$ and in implicit constr.;
- 8: // optimal solution fulfills h with equality, i.e., $a_h^T x = b_h$

Algorithm 1 SeidelLP(\mathcal{H}, d)

- 1: if d = 1 then solve 1-dimensional problem and return; 2: if $\mathcal{H} = \emptyset$ then return x on implicit constraint hyperplane
- 3: choose random constraint $h \in \mathcal{H}$
- 4: $\hat{\mathcal{H}} \leftarrow \mathcal{H} \setminus \{h\}$
- 5: $\hat{x}^* \leftarrow \text{SeidelLP}(\hat{\mathcal{H}}, d)$
- 6: **if** \hat{x}^* = infeasible **then return** infeasible
 - 7: **if** \hat{x}^* fulfills h then return \hat{x}^*
 - 8: // optimal solution fulfills h with equality, i.e., $a_h^T x = b_h$
- 11: $\hat{x}^* \leftarrow \text{SeidelLP}(\hat{\mathcal{H}}, d-1)$
- 12: **if** \hat{x}^* = infeasible **then**

14: **else**

- return infeasible
- 9: solve $a_h^T x = b_h$ for some variable x_ℓ ; 10: eliminate x_{ℓ} in constraints from $\hat{\mathcal{H}}$ and in implicit constr.;

add the value of x_{ℓ} to \hat{x}^* and return the solution

- If d = 1 we can solve the 1-dimensional problem in time $\mathcal{O}(\max\{m, 1\})$.
- If d > 1 and m = 0 we take time O(d) to return d-dimensional vector x.
- ▶ The first recursive call takes time T(m-1,d) for the call plus $\mathcal{O}(d)$ for checking whether the solution fulfills h.
- If we are unlucky and \hat{x}^* does not fulfill h we need time $\mathcal{O}(d(m+1)) = \mathcal{O}(dm)$ to eliminate x_ℓ . Then we make a recursive call that takes time T(m-1,d-1).
- The probability of being unlucky is at most d/m as there are at most d constraints whose removal will decrease the objective function

- If d = 1 we can solve the 1-dimensional problem in time $\mathcal{O}(\max\{m, 1\})$.
- If d > 1 and m = 0 we take time O(d) to return d-dimensional vector x.
- ▶ The first recursive call takes time T(m-1,d) for the call plus O(d) for checking whether the solution fulfills h.
- If we are unlucky and \hat{x}^* does not fulfill h we need time $\mathcal{O}(d(m+1)) = \mathcal{O}(dm)$ to eliminate x_ℓ . Then we make a recursive call that takes time T(m-1,d-1).
- The probability of being unlucky is at most d/m as there are at most d constraints whose removal will decrease the objective function

- If d = 1 we can solve the 1-dimensional problem in time $\mathcal{O}(\max\{m, 1\})$.
- If d > 1 and m = 0 we take time O(d) to return d-dimensional vector x.
- ▶ The first recursive call takes time T(m-1,d) for the call plus O(d) for checking whether the solution fulfills h.
- If we are unlucky and \hat{x}^* does not fulfill h we need time $\mathcal{O}(d(m+1)) = \mathcal{O}(dm)$ to eliminate x_ℓ . Then we make a recursive call that takes time T(m-1,d-1).
- The probability of being unlucky is at most d/m as there are at most d constraints whose removal will decrease the objective function

- If d = 1 we can solve the 1-dimensional problem in time $O(\max\{m, 1\})$.
- If d > 1 and m = 0 we take time O(d) to return d-dimensional vector x.
- ▶ The first recursive call takes time T(m-1,d) for the call plus O(d) for checking whether the solution fulfills h.
- ▶ If we are unlucky and \hat{x}^* does not fulfill h we need time $\mathcal{O}(d(m+1)) = \mathcal{O}(dm)$ to eliminate x_ℓ . Then we make a recursive call that takes time T(m-1,d-1).
- ▶ The probability of being unlucky is at most d/m as there are at most d constraints whose removal will decrease the objective function

- If d = 1 we can solve the 1-dimensional problem in time $O(\max\{m, 1\})$.
- If d > 1 and m = 0 we take time O(d) to return d-dimensional vector x.
- ▶ The first recursive call takes time T(m-1,d) for the call plus O(d) for checking whether the solution fulfills h.
- If we are unlucky and \hat{x}^* does not fulfill h we need time $\mathcal{O}(d(m+1)) = \mathcal{O}(dm)$ to eliminate x_ℓ . Then we make a recursive call that takes time T(m-1,d-1).
- The probability of being unlucky is at most d/m as there are at most d constraints whose removal will decrease the objective function

This gives the recurrence

$$T(m,d) = \begin{cases} \mathcal{O}(\max\{1,m\}) & \text{if } d=1\\ \mathcal{O}(d) & \text{if } d>1 \text{ and } m=0\\ \mathcal{O}(d) + T(m-1,d) +\\ \frac{d}{m}(\mathcal{O}(dm) + T(m-1,d-1)) & \text{otw.} \end{cases}$$

Note that T(m, d) denotes the expected running time.

Let C be the largest constant in the \mathcal{O} -notations.

$$T(m,d) = \begin{cases} C \max\{1,m\} & \text{if } d = 1 \\ Cd & \text{if } d > 1 \text{ and } m = 0 \\ Cd + T(m-1,d) + \\ \frac{d}{m}(Cdm + T(m-1,d-1)) & \text{otw.} \end{cases}$$

Note that T(m, d) denotes the expected running time.

Let $\mathcal C$ be the largest constant in the $\mathcal O$ -notations.

Let C be the largest constant in the \mathcal{O} -notations.

Let \mathcal{C} be the largest constant in the \mathcal{O} -notations.

We show $T(m, d) \le Cf(d) \max\{1, m\}$.

d = 1:

Let C be the largest constant in the \mathcal{O} -notations.

We show $T(m, d) \le Cf(d) \max\{1, m\}$.

d = 1:

T(m, 1)

Let \mathcal{C} be the largest constant in the \mathcal{O} -notations.

We show $T(m, d) \le Cf(d) \max\{1, m\}$.

$$d = 1$$
:

 $T(m,1) \le C \max\{1,m\}$

Let C be the largest constant in the \mathcal{O} -notations.

We show $T(m, d) \le Cf(d) \max\{1, m\}$.

$$d = 1$$
:

 $T(m, 1) \le C \max\{1, m\} \le Cf(1) \max\{1, m\}$

Let C be the largest constant in the \mathcal{O} -notations.

We show $T(m, d) \le Cf(d) \max\{1, m\}$.

$$d = 1$$
:

 $T(m,1) \le C \max\{1,m\} \le Cf(1) \max\{1,m\} \text{ for } f(1) \ge 1$

Let C be the largest constant in the \mathcal{O} -notations.

We show $T(m, d) \leq Cf(d) \max\{1, m\}$.

$$d = 1$$
:

$$T(m,1) \le C \max\{1,m\} \le Cf(1) \max\{1,m\} \text{ for } f(1) \ge 1$$

$$d>1; m=0:$$

 $T(0,d) \leq \mathcal{O}(d)$

Let \mathcal{C} be the largest constant in the \mathcal{O} -notations.

$$d = 1$$
:

$$T(m,1) \le C \max\{1,m\} \le Cf(1) \max\{1,m\} \text{ for } f(1) \ge 1$$

$$d > 1; m = 0:$$

$$T(0,d) \le \mathcal{O}(d) \le Cd$$

Let C be the largest constant in the \mathcal{O} -notations.

$$d = 1$$
:

$$T(m,1) \le C \max\{1,m\} \le Cf(1) \max\{1,m\} \text{ for } f(1) \ge 1$$

$$d > 1; m = 0:$$

$$T(0,d) \le \mathcal{O}(d) \le Cd \le Cf(d) \max\{1,m\}$$

Let C be the largest constant in the \mathcal{O} -notations.

$$d = 1$$
:

$$T(m,1) \le C \max\{1,m\} \le Cf(1) \max\{1,m\} \text{ for } f(1) \ge 1$$

$$d > 1; m = 0$$
:

$$T(0,d) \le \mathcal{O}(d) \le Cd \le Cf(d) \max\{1,m\} \text{ for } f(d) \ge d$$

Let C be the largest constant in the \mathcal{O} -notations.

$$d = 1$$
:

$$T(m,1) \le C \max\{1,m\} \le Cf(1) \max\{1,m\} \text{ for } f(1) \ge 1$$

$$d > 1; m = 0$$
:

$$T(0,d) \le \mathcal{O}(d) \le Cd \le Cf(d) \max\{1,m\} \text{ for } f(d) \ge d$$

$$d > 1; m = 1:$$

$$T(1,d) = \mathcal{O}(d) + T(0,d) + d(\mathcal{O}(d) + T(0,d-1))$$

Let \mathcal{C} be the largest constant in the \mathcal{O} -notations.

We show $T(m, d) \le Cf(d) \max\{1, m\}$.

$$d = 1$$
:

$$T(m,1) \le C \max\{1,m\} \le Cf(1) \max\{1,m\} \text{ for } f(1) \ge 1$$

$$d > 1; m = 0$$
:

$$T(0,d) \le \mathcal{O}(d) \le Cd \le Cf(d) \max\{1,m\} \text{ for } f(d) \ge d$$

d > 1; m = 1:

$$T(1,d) = \mathcal{O}(d) + T(0,d) + d(\mathcal{O}(d) + T(0,d-1))$$

 $\leq Cd + Cd + Cd^2 + dCf(d-1)$

Let C be the largest constant in the \mathcal{O} -notations.

We show $T(m, d) \le Cf(d) \max\{1, m\}$.

$$d = 1$$
:

$$T(m,1) \le C \max\{1,m\} \le Cf(1) \max\{1,m\} \text{ for } f(1) \ge 1$$

$$d > 1; m = 0:$$

$$T(0,d) \le \mathcal{O}(d) \le Cd \le Cf(d) \max\{1,m\} \text{ for } f(d) \ge d$$

d > 1; m = 1:

$$T(1,d) = \mathcal{O}(d) + T(0,d) + d\left(\mathcal{O}(d) + T(0,d-1)\right)$$

$$\leq Cd + Cd + Cd^2 + dCf(d-1)$$

$$\leq Cf(d) \max\{1, m\}$$

Let C be the largest constant in the \mathcal{O} -notations.

We show $T(m, d) \le Cf(d) \max\{1, m\}$.

$$d = 1$$
:

$$T(m,1) \le C \max\{1,m\} \le Cf(1) \max\{1,m\} \text{ for } f(1) \ge 1$$

$$d > 1; m = 0$$
:

$$T(0,d) \le \mathcal{O}(d) \le Cd \le Cf(d) \max\{1,m\} \text{ for } f(d) \ge d$$

d > 1; m = 1:

$$T(1,d) = O(d) + T(0,d) + d(O(d) + T(0,d-1))$$

$$\leq Cd + Cd + Cd^{2} + dCf(d-1)$$

$$\leq Cf(d) \max\{1, m\} \text{ for } f(d) \geq 3d^{2} + df(d-1)$$

```
d > 1; m > 1:
(by induction hypothesis statm. true for d' < d, m' \ge 0; and for d' = d, m' < m)
```

d > 1; m > 1:

$$T(m,d) = \mathcal{O}(d) + T(m-1,d) + \frac{d}{m} \left(\mathcal{O}(dm) + T(m-1,d-1) \right)$$

d > 1; m > 1:

$$T(m,d) = \mathcal{O}(d) + T(m-1,d) + \frac{d}{m} \Big(\mathcal{O}(dm) + T(m-1,d-1) \Big)$$

$$\leq Cd + Cf(d)(m-1) + Cd^2 + \frac{d}{m} Cf(d-1)(m-1)$$

d > 1; m > 1:

$$T(m,d) = \mathcal{O}(d) + T(m-1,d) + \frac{d}{m} \Big(\mathcal{O}(dm) + T(m-1,d-1) \Big)$$

$$\leq Cd + Cf(d)(m-1) + Cd^2 + \frac{d}{m}Cf(d-1)(m-1)$$

$$\leq 2Cd^2 + Cf(d)(m-1) + dCf(d-1)$$

d > 1: m > 1:

$$T(m,d) = \mathcal{O}(d) + T(m-1,d) + \frac{d}{m} \Big(\mathcal{O}(dm) + T(m-1,d-1) \Big)$$

$$\leq Cd + Cf(d)(m-1) + Cd^2 + \frac{d}{m}Cf(d-1)(m-1)$$

$$\leq 2Cd^2 + Cf(d)(m-1) + dCf(d-1)$$

$$\leq Cf(d)m$$

d > 1: m > 1:

$$\begin{split} T(m,d) &= \mathcal{O}(d) + T(m-1,d) + \frac{d}{m} \Big(\mathcal{O}(dm) + T(m-1,d-1) \Big) \\ &\leq Cd + Cf(d)(m-1) + Cd^2 + \frac{d}{m} Cf(d-1)(m-1) \\ &\leq 2Cd^2 + Cf(d)(m-1) + dCf(d-1) \\ &\leq Cf(d)m \end{split}$$

if
$$f(d) \ge df(d-1) + 2d^2$$
.

▶ Define $f(1) = 3 \cdot 1^2$ and $f(d) = df(d-1) + 3d^2$ for d > 1.

▶ Define $f(1) = 3 \cdot 1^2$ and $f(d) = df(d-1) + 3d^2$ for d > 1.

Then

f(d)

▶ Define $f(1) = 3 \cdot 1^2$ and $f(d) = df(d-1) + 3d^2$ for d > 1.

$$f(d) = 3d^2 + df(d-1)$$

▶ Define $f(1) = 3 \cdot 1^2$ and $f(d) = df(d-1) + 3d^2$ for d > 1.

$$f(d) = 3d^{2} + df(d-1)$$
$$= 3d^{2} + d\left[3(d-1)^{2} + (d-1)f(d-2)\right]$$

▶ Define $f(1) = 3 \cdot 1^2$ and $f(d) = df(d-1) + 3d^2$ for d > 1.

$$f(d) = 3d^{2} + df(d-1)$$

$$= 3d^{2} + d\left[3(d-1)^{2} + (d-1)f(d-2)\right]$$

$$= 3d^{2} + d\left[3(d-1)^{2} + (d-1)\left[3(d-2)^{2} + (d-2)f(d-3)\right]\right]$$

▶ Define $f(1) = 3 \cdot 1^2$ and $f(d) = df(d-1) + 3d^2$ for d > 1.

$$f(d) = 3d^{2} + df(d-1)$$

$$= 3d^{2} + d\left[3(d-1)^{2} + (d-1)f(d-2)\right]$$

$$= 3d^{2} + d\left[3(d-1)^{2} + (d-1)\left[3(d-2)^{2} + (d-2)f(d-3)\right]\right]$$

$$= 3d^{2} + 3d(d-1)^{2} + 3d(d-1)(d-2)^{2} + \dots$$

$$+ 3d(d-1)(d-2) \cdot \dots \cdot 4 \cdot 3 \cdot 2 \cdot 1^{2}$$

8 Seidels LP-algorithm

▶ Define $f(1) = 3 \cdot 1^2$ and $f(d) = df(d-1) + 3d^2$ for d > 1.

Then

$$f(d) = 3d^{2} + df(d-1)$$

$$= 3d^{2} + d\left[3(d-1)^{2} + (d-1)f(d-2)\right]$$

$$= 3d^{2} + d\left[3(d-1)^{2} + (d-1)\left[3(d-2)^{2} + (d-2)f(d-3)\right]\right]$$

$$= 3d^{2} + 3d(d-1)^{2} + 3d(d-1)(d-2)^{2} + \dots$$

$$+ 3d(d-1)(d-2) \cdot \dots \cdot 4 \cdot 3 \cdot 2 \cdot 1^{2}$$

$$= 3d! \left(\frac{d^{2}}{d!} + \frac{(d-1)^{2}}{(d-1)!} + \frac{(d-2)^{2}}{(d-2)!} + \dots\right)$$

8 Seidels LP-algorithm

▶ Define $f(1) = 3 \cdot 1^2$ and $f(d) = df(d-1) + 3d^2$ for d > 1.

Then

$$f(d) = 3d^{2} + df(d-1)$$

$$= 3d^{2} + d\left[3(d-1)^{2} + (d-1)f(d-2)\right]$$

$$= 3d^{2} + d\left[3(d-1)^{2} + (d-1)\left[3(d-2)^{2} + (d-2)f(d-3)\right]\right]$$

$$= 3d^{2} + 3d(d-1)^{2} + 3d(d-1)(d-2)^{2} + \dots$$

$$+ 3d(d-1)(d-2) \cdot \dots \cdot 4 \cdot 3 \cdot 2 \cdot 1^{2}$$

$$= 3d! \left(\frac{d^{2}}{d!} + \frac{(d-1)^{2}}{(d-1)!} + \frac{(d-2)^{2}}{(d-2)!} + \dots\right)$$

$$= \mathcal{O}(d!)$$

8 Seidels LP-algorithm

▶ Define $f(1) = 3 \cdot 1^2$ and $f(d) = df(d-1) + 3d^2$ for d > 1.

Then

$$f(d) = 3d^{2} + df(d-1)$$

$$= 3d^{2} + d\left[3(d-1)^{2} + (d-1)f(d-2)\right]$$

$$= 3d^{2} + d\left[3(d-1)^{2} + (d-1)\left[3(d-2)^{2} + (d-2)f(d-3)\right]\right]$$

$$= 3d^{2} + 3d(d-1)^{2} + 3d(d-1)(d-2)^{2} + \dots$$

$$+ 3d(d-1)(d-2) \cdot \dots \cdot 4 \cdot 3 \cdot 2 \cdot 1^{2}$$

$$= 3d! \left(\frac{d^{2}}{d!} + \frac{(d-1)^{2}}{(d-1)!} + \frac{(d-2)^{2}}{(d-2)!} + \dots\right)$$

$$= \mathcal{O}(d!)$$

since $\sum_{i\geq 1}\frac{i^2}{i!}$ is a constant.

Complexity

LP Feasibility Problem (LP feasibility A)

Given $A \in \mathbb{Z}^{m \times n}$, $b \in \mathbb{Z}^m$. Does there exist $x \in \mathbb{R}^n$ with $Ax \le b$, $x \ge 0$?

LP Feasibility Problem (LP feasibility B)

Given $A \in \mathbb{Z}^{m \times n}$, $b \in \mathbb{Z}^m$. Find $x \in \mathbb{R}^n$ with $Ax \le b$, $x \ge 0$!

LP Optimization A

Given $A \in \mathbb{Z}^{m \times n}$, $b \in \mathbb{Z}^m$, $c \in \mathbb{Z}^n$. What is the maximum value of $c^T x$ for a feasible point $x \in \mathbb{R}^n$?

LP Optimization B

Given $A \in \mathbb{Z}^{m \times n}$, $b \in \mathbb{Z}^m$, $c \in \mathbb{Z}^n$. Return feasible point $x \in \mathbb{R}^n$ with maximum value of $c^T x$?

Input size

▶ The number of bits to represent a number $a \in \mathbb{Z}$ is

$$\lceil \log_2(|a|) \rceil + 1$$

$$\langle M
angle := \sum_{i,j} \lceil \log_2(|m_{ij}|) + 1 \rceil$$

- In the following we assume that input matrices are encoded in a standard way, where each number is encoded in binary and then suitable separators are added in order to separate distinct number from each other.
- ▶ Then the input length is $L = \Theta(\langle A \rangle + \langle b \rangle)$

Input size

▶ The number of bits to represent a number $a \in \mathbb{Z}$ is

$$\lceil \log_2(|a|) \rceil + 1$$

$$\langle M \rangle := \sum_{i,j} \lceil \log_2(|m_{ij}|) + 1 \rceil$$

- In the following we assume that input matrices are encoded in a standard way, where each number is encoded in binary and then suitable separators are added in order to separate distinct number from each other.
- ▶ Then the input length is $L = \Theta(\langle A \rangle + \langle b \rangle)$.

Input size

▶ The number of bits to represent a number $a \in \mathbb{Z}$ is

$$\lceil \log_2(|a|) \rceil + 1$$

$$\langle M \rangle := \sum_{i,j} \lceil \log_2(|m_{ij}|) + 1 \rceil$$

- In the following we assume that input matrices are encoded in a standard way, where each number is encoded in binary and then suitable separators are added in order to separate distinct number from each other.
- ▶ Then the input length is $L = \Theta(\langle A \rangle + \langle b \rangle)$.

Input size

▶ The number of bits to represent a number $a \in \mathbb{Z}$ is

$$\lceil \log_2(|a|) \rceil + 1$$

$$\langle M \rangle := \sum_{i,j} \lceil \log_2(|m_{ij}|) + 1 \rceil$$

- ▶ In the following we assume that input matrices are encoded in a standard way, where each number is encoded in binary and then suitable separators are added in order to separate distinct number from each other.
- ▶ Then the input length is $L = \Theta(\langle A \rangle + \langle b \rangle)$.

- In the following we sometimes refer to $L := \langle A \rangle + \langle b \rangle$ as the input size (even though the real input size is something in $\Theta(\langle A \rangle + \langle b \rangle)$).
- Sometimes we may also refer to $L := \langle A \rangle + \langle b \rangle + n \log_2 n$ as the input size. Note that $n \log_2 n = \Theta(\langle A \rangle + \langle b \rangle)$.
- ▶ In order to show that LP-decision is in NP we show that if there is a solution *x* then there exists a small solution for which feasibility can be verified in polynomial time (polynomial in *L*).

Suppose that $\bar{A}x = b$; $x \ge 0$ is feasible.

Then there exists a basic feasible solution. This means a set B of basic variables such that

$$x_B = \bar{A}_B^{-1} b$$

and all other entries in x are 0.

Suppose that $\bar{A}x = b$; $x \ge 0$ is feasible.

Then there exists a basic feasible solution. This means a set B of basic variables such that

$$x_B = \bar{A}_B^{-1} b$$

and all other entries in x are 0.

Size of a Basic Feasible Solution

- A: original input matrix
- \blacktriangleright \bar{A} : transformation of A into standard form
- lacksquare $ar{A}_B$: submatrix of $ar{A}$ corresponding to basis B

Lemma 47

Let $\bar{A}_B \in \mathbb{Z}^{m \times m}$ and $b \in \mathbb{Z}^m$. Define $L = \langle A \rangle + \langle b \rangle + n \log_2 n$. Then a solution to $\bar{A}_B x_B = b$ has rational components x_j of the form $\frac{D_j}{D}$, where $|D_j| \leq 2^L$ and $|D| \leq 2^L$.

Proof

Cramers rules says that we can compute x_j as

$$x_j = \frac{\det(\bar{A}_B^j)}{\det(\bar{A}_B)}$$

where \bar{A}_B^J is the matrix obtained from \bar{A}_B by replacing the j-th column by the vector b.

Size of a Basic Feasible Solution

- ► A: original input matrix
- \blacktriangleright \bar{A} : transformation of A into standard form
- $ightharpoonup ar{A}_B$: submatrix of $ar{A}$ corresponding to basis B

Lemma 47

Let $\bar{A}_B \in \mathbb{Z}^{m \times m}$ and $b \in \mathbb{Z}^m$. Define $L = \langle A \rangle + \langle b \rangle + n \log_2 n$. Then a solution to $\bar{A}_B x_B = b$ has rational components x_j of the form $\frac{D_j}{D}$, where $|D_j| \leq 2^L$ and $|D| \leq 2^L$.

Proof:

Cramers rules says that we can compute x_j as

$$x_j = \frac{\det(\bar{A}_B^j)}{\det(\bar{A}_B)}$$

where \bar{A}_B^j is the matrix obtained from \bar{A}_B by replacing the j-th column by the vector b.

Let
$$X = \bar{A}_B$$
. Then

 $|\det(X)|$

Let
$$X = \bar{A}_B$$
. Then

$$|\det(X)| = |\det(\bar{X})|$$

Let
$$X = \bar{A}_B$$
. Then

$$|\det(X)| = |\det(\bar{X})|$$

$$= \left| \sum_{\pi \in S_{\tilde{n}}} \operatorname{sgn}(\pi) \prod_{1 \le i \le \tilde{n}} \bar{X}_{i\pi(i)} \right|$$

Let
$$X = \bar{A}_B$$
. Then

$$\begin{aligned} |\det(X)| &= |\det(\bar{X})| \\ &= \left| \sum_{\pi \in S_{\bar{n}}} \operatorname{sgn}(\pi) \prod_{1 \le i \le \bar{n}} \bar{X}_{i\pi(i)} \right| \\ &\le \sum_{\pi \in S_{\bar{n}}} \prod_{1 \le i \le \bar{n}} |\bar{X}_{i\pi(i)}| \end{aligned}$$

Let
$$X = \bar{A}_B$$
. Then

$$\begin{aligned} |\det(X)| &= |\det(\bar{X})| \\ &= \left| \sum_{\pi \in S_{\tilde{n}}} \operatorname{sgn}(\pi) \prod_{1 \le i \le \tilde{n}} \bar{X}_{i\pi(i)} \right| \\ &\le \sum_{\pi \in S_{\tilde{n}}} \prod_{1 \le i \le \tilde{n}} |\bar{X}_{i\pi(i)}| \\ &\le n! \cdot 2^{\langle A \rangle + \langle b \rangle} \end{aligned}$$

Let
$$X = \bar{A}_B$$
. Then

$$\begin{aligned} |\det(X)| &= |\det(\bar{X})| \\ &= \left| \sum_{\pi \in S_{\tilde{n}}} \operatorname{sgn}(\pi) \prod_{1 \le i \le \tilde{n}} \bar{X}_{i\pi(i)} \right| \\ &\le \sum_{\pi \in S_{\tilde{n}}} \prod_{1 \le i \le \tilde{n}} |\bar{X}_{i\pi(i)}| \\ &\le n! \cdot 2^{\langle A \rangle + \langle b \rangle} \le 2^{L} \ . \end{aligned}$$

Let
$$X = \bar{A}_B$$
. Then

$$\begin{aligned} |\text{det}(X)| &= |\text{det}(\bar{X})| \\ &= \left| \sum_{\pi \in S_{\tilde{n}}} \text{sgn}(\pi) \prod_{1 \leq i \leq \tilde{n}} \bar{X}_{i\pi(i)} \right| \\ &\leq \sum_{\pi \in S_{\tilde{n}}} \prod_{1 \leq i \leq \tilde{n}} |\bar{X}_{i\pi(i)}| \\ &\leq n! \cdot 2^{\langle A \rangle + \langle b \rangle} \leq 2^{L} \ . \end{aligned}$$

Here \bar{X} is an $\tilde{n} \times \tilde{n}$ submatrix of A with $\tilde{n} \leq n$.

Let
$$X = \bar{A}_B$$
. Then

$$\begin{aligned} |\text{det}(X)| &= |\text{det}(\bar{X})| \\ &= \left| \sum_{\pi \in S_{\tilde{n}}} \text{sgn}(\pi) \prod_{1 \leq i \leq \tilde{n}} \bar{X}_{i\pi(i)} \right| \\ &\leq \sum_{\pi \in S_{\tilde{n}}} \prod_{1 \leq i \leq \tilde{n}} |\bar{X}_{i\pi(i)}| \\ &\leq n! \cdot 2^{\langle A \rangle + \langle b \rangle} \leq 2^{L} \ . \end{aligned}$$

Here \bar{X} is an $\tilde{n} \times \tilde{n}$ submatrix of A with $\tilde{n} \leq n$.

Analogously for $\det(A_R^j)$.

Given an LP $\max\{c^Tx \mid Ax \leq b; x \geq 0\}$ do a binary search for the optimum solution

(Add constraint $c^T x \ge M$). Then checking for feasibility shows whether optimum solution is larger or smaller than M).

If the LP is feasible then the binary search finishes in at most

$$\log_2\left(\frac{2n2^{2L'}}{1/2^{L'}}\right) = \mathcal{O}(L') ,$$

as the range of the search is at most $-n2^{2L'},\ldots,n2^{2L'}$ and the distance between two adjacent values is at least $\frac{1}{\det(A)} \geq \frac{1}{2^{L'}}$.

Given an LP $\max\{c^Tx \mid Ax \leq b; x \geq 0\}$ do a binary search for the optimum solution

(Add constraint $c^T x \ge M$). Then checking for feasibility shows whether optimum solution is larger or smaller than M).

If the LP is feasible then the binary search finishes in at most

$$\log_2\left(\frac{2n2^{2L'}}{1/2^{L'}}\right) = \mathcal{O}(L') ,$$

as the range of the search is at most $-n2^{2L'},\ldots,n2^{2L'}$ and the distance between two adjacent values is at least $\frac{1}{\det(A)} \geq \frac{1}{2^{L'}}$.

Given an LP $\max\{c^Tx \mid Ax \leq b; x \geq 0\}$ do a binary search for the optimum solution

(Add constraint $c^T x \ge M$). Then checking for feasibility shows whether optimum solution is larger or smaller than M).

If the LP is feasible then the binary search finishes in at most

$$\log_2\left(\frac{2n2^{2L'}}{1/2^{L'}}\right) = \mathcal{O}(L') ,$$

as the range of the search is at most $-n2^{2L'},\ldots,n2^{2L'}$ and the distance between two adjacent values is at least $\frac{1}{\det(A)} \geq \frac{1}{2^{L'}}$.

Given an LP $\max\{c^Tx \mid Ax \leq b; x \geq 0\}$ do a binary search for the optimum solution

(Add constraint $c^T x \ge M$). Then checking for feasibility shows whether optimum solution is larger or smaller than M).

If the LP is feasible then the binary search finishes in at most

$$\log_2\left(\frac{2n2^{2L'}}{1/2^{L'}}\right) = \mathcal{O}(L') ,$$

as the range of the search is at most $-n2^{2L'},\ldots,n2^{2L'}$ and the distance between two adjacent values is at least $\frac{1}{\det(A)} \geq \frac{1}{2^{L'}}$.

Given an LP $\max\{c^Tx \mid Ax \leq b; x \geq 0\}$ do a binary search for the optimum solution

(Add constraint $c^T x \ge M$). Then checking for feasibility shows whether optimum solution is larger or smaller than M).

If the LP is feasible then the binary search finishes in at most

$$\log_2\left(\frac{2n2^{2L'}}{1/2^{L'}}\right) = \mathcal{O}(L') ,$$

as the range of the search is at most $-n2^{2L'}, \ldots, n2^{2L'}$ and the distance between two adjacent values is at least $\frac{1}{\det(A)} \ge \frac{1}{2^{L'}}$.

How do we detect whether the LP is unbounded?

Let $M_{\text{max}} = n2^{2L'}$ be an upper bound on the objective value of a basic feasible solution.

We can add a constraint $c^T x \ge M_{\text{max}} + 1$ and check for feasibility.

How do we detect whether the LP is unbounded?

Let $M_{\text{max}} = n2^{2L'}$ be an upper bound on the objective value of a basic feasible solution.

We can add a constraint $c^T x \ge M_{\text{max}} + 1$ and check for feasibility.

How do we detect whether the LP is unbounded?

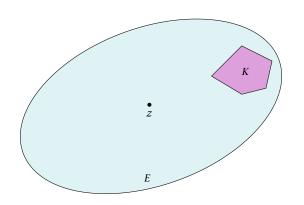
Let $M_{\text{max}} = n2^{2L'}$ be an upper bound on the objective value of a basic feasible solution.

We can add a constraint $c^T x \ge M_{\text{max}} + 1$ and check for feasibility.

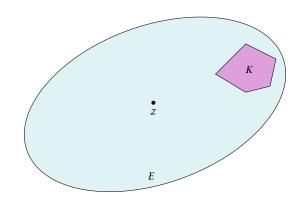
Let *K* be a convex set.



- Let *K* be a convex set.
- Maintain ellipsoid E that is guaranteed to contain K provided that K is non-empty.



- Let *K* be a convex set.
- Maintain ellipsoid E that is guaranteed to contain K provided that K is non-empty.
- ▶ If center $z \in K$ STOP.



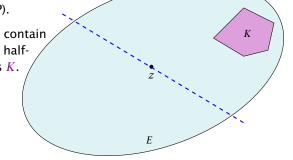
- Let *K* be a convex set.
- Maintain ellipsoid E that is guaranteed to contain K provided that K is non-empty.
- ▶ If center $z \in K$ STOP.

Otw. find a hyperplane separating K from z (e.g. a violated constraint in the LP).

- Let K be a convex set.
- Maintain ellipsoid *E* that is guaranteed to contain *K* provided that *K* is non-empty.
- ▶ If center $z \in K$ STOP.

Otw. find a hyperplane separating K from z (e.g. a violated constraint in the LP).

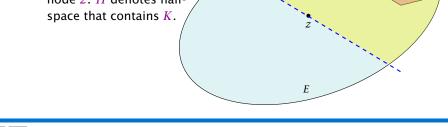
Shift hyperplane to contain node z. H denotes halfspace that contains K.



- Let K be a convex set.
- Maintain ellipsoid *E* that is guaranteed to contain *K* provided that *K* is non-empty.
- ▶ If center $z \in K$ STOP.

Otw. find a hyperplane separating K from z (e.g. a violated constraint in the LP).

Shift hyperplane to contain node z. H denotes halfspace that contains K.



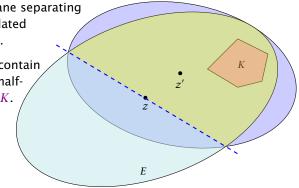
Ellipsoid Method

- Let K be a convex set.
- Maintain ellipsoid *E* that is guaranteed to contain *K* provided that *K* is non-empty.
- ▶ If center $z \in K$ STOP.

Otw. find a hyperplane separating K from z (e.g. a violated constraint in the LP).

Shift hyperplane to contain node z. H denotes halfspace that contains K.

Compute (smallest) ellipsoid E' that contains $E \cap H$.

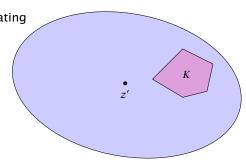


Ellipsoid Method

- Let K be a convex set.
- Maintain ellipsoid *E* that is guaranteed to contain *K* provided that *K* is non-empty.
- ▶ If center $z \in K$ STOP.

Otw. find a hyperplane separating K from z (e.g. a violated constraint in the LP).

- Shift hyperplane to contain node z. H denotes halfspace that contains K.
- Compute (smallest) ellipsoid E' that contains $E \cap H$.

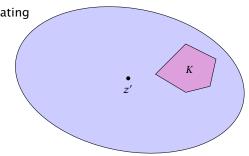


Ellipsoid Method

- Let K be a convex set.
- Maintain ellipsoid *E* that is guaranteed to contain *K* provided that *K* is non-empty.
- ▶ If center $z \in K$ STOP.

Otw. find a hyperplane separating K from z (e.g. a violated constraint in the LP).

- Shift hyperplane to contain node z. H denotes halfspace that contains K.
- Compute (smallest) ellipsoid E' that contains $E \cap H$.
- REPEAT



Issues/Questions:

- How do you choose the first Ellipsoid? What is its volume?
- How do you measure progress? By how much does the volume decrease in each iteration?
- When can you stop? What is the minimum volume of a non-empty polytop?

A mapping $f: \mathbb{R}^n \to \mathbb{R}^n$ with f(x) = Lx + t, where L is an invertible matrix is called an affine transformation.

A ball in \mathbb{R}^n with center c and radius r is given by

$$B(c,r) = \{x \mid (x-c)^T (x-c) \le r^2\}$$
$$= \{x \mid \sum_i (x-c)_i^2 / r^2 \le 1\}$$

B(0,1) is called the unit ball.

An affine transformation of the unit ball is called an ellipsoid.

From
$$f(x) = Lx + t$$
 follows $x = L^{-1}(f(x) - t)$.

170/249

From
$$f(x) = Lx + t$$
 follows $x = L^{-1}(f(x) - t)$.

From
$$f(x) = Lx + t$$
 follows $x = L^{-1}(f(x) - t)$.

$$f(B(0,1)) = \{f(x) \mid x \in B(0,1)\}$$

From
$$f(x) = Lx + t$$
 follows $x = L^{-1}(f(x) - t)$.

$$f(B(0,1)) = \{ f(x) \mid x \in B(0,1) \}$$
$$= \{ y \in \mathbb{R}^n \mid L^{-1}(y-t) \in B(0,1) \}$$

From
$$f(x) = Lx + t$$
 follows $x = L^{-1}(f(x) - t)$.

$$\begin{split} f(B(0,1)) &= \{ f(x) \mid x \in B(0,1) \} \\ &= \{ y \in \mathbb{R}^n \mid L^{-1}(y-t) \in B(0,1) \} \\ &= \{ y \in \mathbb{R}^n \mid (y-t)^T L^{-1}^T L^{-1}(y-t) \le 1 \} \end{split}$$

From
$$f(x) = Lx + t$$
 follows $x = L^{-1}(f(x) - t)$.

$$f(B(0,1)) = \{f(x) \mid x \in B(0,1)\}$$

$$= \{y \in \mathbb{R}^n \mid L^{-1}(y-t) \in B(0,1)\}$$

$$= \{y \in \mathbb{R}^n \mid (y-t)^T L^{-1}^T L^{-1}(y-t) \le 1\}$$

$$= \{y \in \mathbb{R}^n \mid (y-t)^T Q^{-1}(y-t) \le 1\}$$

An affine transformation of the unit ball is called an ellipsoid.

From
$$f(x) = Lx + t$$
 follows $x = L^{-1}(f(x) - t)$.

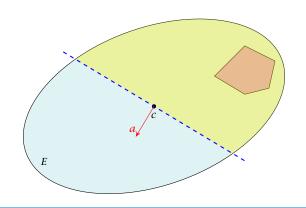
$$f(B(0,1)) = \{ f(x) \mid x \in B(0,1) \}$$

$$= \{ y \in \mathbb{R}^n \mid L^{-1}(y-t) \in B(0,1) \}$$

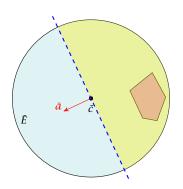
$$= \{ y \in \mathbb{R}^n \mid (y-t)^T L^{-1}^T L^{-1}(y-t) \le 1 \}$$

$$= \{ y \in \mathbb{R}^n \mid (y-t)^T Q^{-1}(y-t) \le 1 \}$$

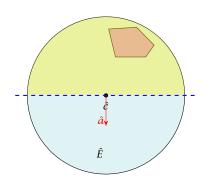
where $Q = LL^T$ is an invertible matrix.



▶ Use f^{-1} (recall that f = Lx + t is the affine transformation of the unit ball) to rotate/distort the ellipsoid (back) into the unit ball.

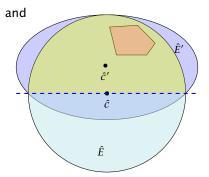


- ▶ Use f^{-1} (recall that f = Lx + t is the affine transformation of the unit ball) to rotate/distort the ellipsoid (back) into the unit ball.
- Use a rotation R^{-1} to rotate the unit ball such that the normal vector of the halfspace is parallel to e_1 .

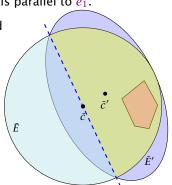


- ▶ Use f^{-1} (recall that f = Lx + t is the affine transformation of the unit ball) to rotate/distort the ellipsoid (back) into the unit ball.
- Use a rotation R^{-1} to rotate the unit ball such that the normal vector of the halfspace is parallel to e_1 .

Compute the new center \hat{c}' and the new matrix \hat{Q}' for this simplified setting.



- ▶ Use f^{-1} (recall that f = Lx + t is the affine transformation of the unit ball) to rotate/distort the ellipsoid (back) into the unit ball.
- Use a rotation R^{-1} to rotate the unit ball such that the normal vector of the halfspace is parallel to e_1 .
- Compute the new center \hat{c}' and the new matrix \hat{Q}' for this simplified setting.
- Use the transformations R and f to get the new center c' and the new matrix Q' for the original ellipsoid E.

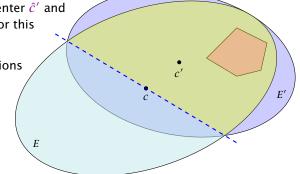


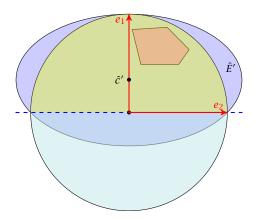
▶ Use f^{-1} (recall that f = Lx + t is the affine transformation of the unit ball) to rotate/distort the ellipsoid (back) into the unit ball.

Use a rotation R^{-1} to rotate the unit ball such that the normal vector of the halfspace is parallel to e_1 .

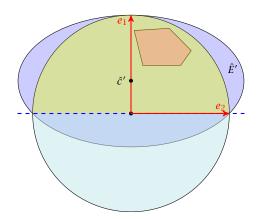
Compute the new center \hat{c}' and the new matrix \hat{Q}' for this simplified setting.

Use the transformations R and f to get the new center c' and the new matrix Q' for the original ellipsoid E.





- ▶ The new center lies on axis x_1 . Hence, $\hat{c}' = te_1$ for t > 0.
- ► The vectors $e_1, e_2,...$ have to fulfill the ellipsoid constraint with equality. Hence $(e_i \hat{c}')^T \hat{O}'^{-1} (e_i \hat{c}') = 1$.



- ▶ The new center lies on axis x_1 . Hence, $\hat{c}' = te_1$ for t > 0.
- ► The vectors $e_1, e_2,...$ have to fulfill the ellipsoid constraint with equality. Hence $(e_i \hat{c}')^T \hat{Q}'^{-1} (e_i \hat{c}') = 1$.

- ▶ To obtain the matrix \hat{Q}'^{-1} for our ellipsoid \hat{E}' note that \hat{E}' is axis-parallel.
- Let a denote the radius along the x_1 -axis and let b denote the (common) radius for the other axes.
- ▶ The matrix

$$\hat{L}' = \begin{pmatrix} a & 0 & \dots & 0 \\ 0 & b & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & b \end{pmatrix}$$

maps the unit ball (via function $\hat{f}'(x) = \hat{L}'x$) to an axis-parallel ellipsoid with radius a in direction x_1 and b in all other directions.

- ▶ To obtain the matrix \hat{Q}'^{-1} for our ellipsoid \hat{E}' note that \hat{E}' is axis-parallel.
- Let a denote the radius along the x_1 -axis and let b denote the (common) radius for the other axes.
- ► The matrix

$$\hat{L}' = \left(\begin{array}{cccc} a & 0 & \dots & 0 \\ 0 & b & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & b \end{array} \right)$$

maps the unit ball (via function $\hat{f}'(x) = \hat{L}'x$) to an axis-parallel ellipsoid with radius a in direction x_1 and b in all other directions.

- ▶ To obtain the matrix \hat{Q}'^{-1} for our ellipsoid \hat{E}' note that \hat{E}' is axis-parallel.
- Let a denote the radius along the x_1 -axis and let b denote the (common) radius for the other axes.
- The matrix

$$\hat{L}' = \begin{pmatrix} a & 0 & \dots & 0 \\ 0 & b & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & b \end{pmatrix}$$

maps the unit ball (via function $\hat{f}'(x) = \hat{L}'x$) to an axis-parallel ellipsoid with radius a in direction x_1 and b in all other directions.

As $\hat{Q}' = \hat{L}'\hat{L}'^t$ the matrix \hat{Q}'^{-1} is of the form

$$\hat{Q}'^{-1} = \begin{pmatrix} \frac{1}{a^2} & 0 & \dots & 0\\ 0 & \frac{1}{b^2} & \ddots & \vdots\\ \vdots & \ddots & \ddots & 0\\ 0 & \dots & 0 & \frac{1}{b^2} \end{pmatrix}$$

 $(e_1 - \hat{c}')^T \hat{Q}'^{-1} (e_1 - \hat{c}') = 1$ gives

$$\begin{pmatrix} 1-t \\ 0 \\ \vdots \\ 0 \end{pmatrix}^{T} \cdot \begin{pmatrix} \frac{1}{a^{2}} & 0 & \cdots & 0 \\ 0 & \frac{1}{b^{2}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \frac{1}{b^{2}} \end{pmatrix} \cdot \begin{pmatrix} 1-t \\ 0 \\ \vdots \\ 0 \end{pmatrix} = 1$$

► This gives $(1 - t)^2 = a^2$.

For $i \neq 1$ the equation $(e_i - \hat{c}')^T \hat{Q}'^{-1} (e_i - \hat{c}') = 1$ looks like (here i = 2)

$$\begin{pmatrix} -t \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}^{T} \cdot \begin{pmatrix} \frac{1}{a^{2}} & 0 & \dots & 0 \\ 0 & \frac{1}{b^{2}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \frac{1}{b^{2}} \end{pmatrix} \cdot \begin{pmatrix} -t \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = 1$$

This gives $\frac{t^2}{a^2} + \frac{1}{b^2} = 1$, and hence

$$\frac{1}{b^2} = 1 - \frac{t^2}{a^2}$$

For $i \neq 1$ the equation $(e_i - \hat{c}')^T \hat{Q}'^{-1} (e_i - \hat{c}') = 1$ looks like (here i = 2)

$$\begin{pmatrix} -t \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}^{T} \cdot \begin{pmatrix} \frac{1}{a^{2}} & 0 & \dots & 0 \\ 0 & \frac{1}{b^{2}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \frac{1}{b^{2}} \end{pmatrix} \cdot \begin{pmatrix} -t \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = 1$$

► This gives $\frac{t^2}{a^2} + \frac{1}{b^2} = 1$, and hence

$$\frac{1}{b^2} = 1 - \frac{t^2}{a^2} = 1 - \frac{t^2}{(1-t)^2}$$

For $i \neq 1$ the equation $(e_i - \hat{c}')^T \hat{Q}'^{-1} (e_i - \hat{c}') = 1$ looks like (here i = 2)

$$\begin{pmatrix} -t \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}^{T} \cdot \begin{pmatrix} \frac{1}{a^{2}} & 0 & \dots & 0 \\ 0 & \frac{1}{b^{2}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \frac{1}{b^{2}} \end{pmatrix} \cdot \begin{pmatrix} -t \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = 1$$

This gives $\frac{t^2}{a^2} + \frac{1}{b^2} = 1$, and hence

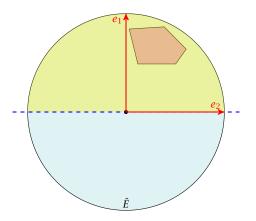
$$\frac{1}{b^2} = 1 - \frac{t^2}{a^2} = 1 - \frac{t^2}{(1-t)^2} = \frac{1-2t}{(1-t)^2}$$

Summary

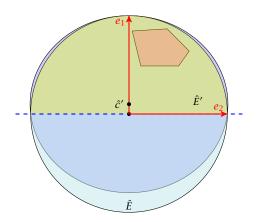
So far we have

$$a = 1 - t$$
 and $b = \frac{1 - t}{\sqrt{1 - 2t}}$

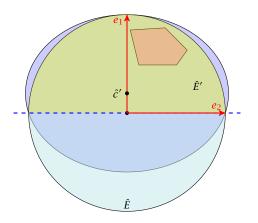
We still have many choices for t:



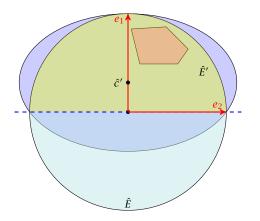
We still have many choices for t:



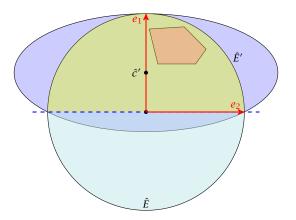
We still have many choices for t:



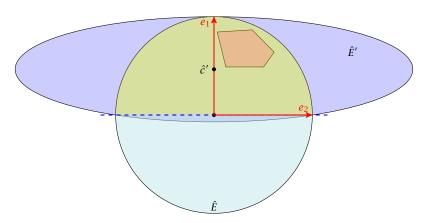
We still have many choices for t:



We still have many choices for t:

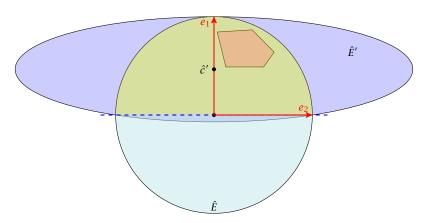


We still have many choices for t:



Choose t such that the volume of \hat{E}' is minimal!!!

We still have many choices for t:



Choose t such that the volume of \hat{E}' is minimal!!!

We want to choose t such that the volume of \hat{E}' is minimal.

Lemma 51

Let L be an affine transformation and $K\subseteq\mathbb{R}^n.$ Then

 $vol(L(K)) = |det(L)| \cdot vol(K)$.

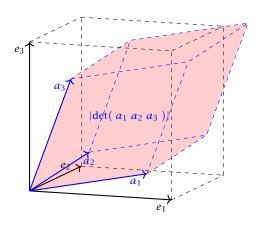
We want to choose t such that the volume of \hat{E}' is minimal.

Lemma 51

Let L be an affine transformation and $K \subseteq \mathbb{R}^n$. Then

$$vol(L(K)) = |det(L)| \cdot vol(K)$$
.

n-dimensional volume



• We want to choose t such that the volume of \hat{E}' is minimal.

$$\operatorname{vol}(\hat{E}') = \operatorname{vol}(B(0,1)) \cdot |\det(\hat{L}')| ,$$

Recall that

$$\hat{L}' = \left(\begin{array}{cccc} a & 0 & \dots & 0 \\ 0 & b & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & b \end{array}\right)$$

Note that *a* and *b* in the above equations depend on *t*, by the previous equations.

• We want to choose t such that the volume of \hat{E}' is minimal.

$$\operatorname{vol}(\hat{E}') = \operatorname{vol}(B(0,1)) \cdot |\det(\hat{L}')| ,$$

Recall that

$$\hat{L}' = \begin{pmatrix} a & 0 & \dots & 0 \\ 0 & b & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & b \end{pmatrix}$$

Note that *a* and *b* in the above equations depend on *t*, by the previous equations.

• We want to choose t such that the volume of \hat{E}' is minimal.

$$\operatorname{vol}(\hat{E}') = \operatorname{vol}(B(0,1)) \cdot |\det(\hat{L}')| ,$$

Recall that

$$\hat{L}' = \begin{pmatrix} a & 0 & \dots & 0 \\ 0 & b & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & b \end{pmatrix}$$

Note that *a* and *b* in the above equations depend on *t*, by the previous equations.

 $\operatorname{vol}(\hat{E}')$

$$\operatorname{vol}(\hat{E}') = \operatorname{vol}(B(0,1)) \cdot |\det(\hat{L}')|$$

$$\operatorname{vol}(\hat{E}') = \operatorname{vol}(B(0,1)) \cdot |\det(\hat{L}')|$$
$$= \operatorname{vol}(B(0,1)) \cdot ab^{n-1}$$

$$\begin{aligned} \operatorname{vol}(\hat{E}') &= \operatorname{vol}(B(0,1)) \cdot |\det(\hat{L}')| \\ &= \operatorname{vol}(B(0,1)) \cdot ab^{n-1} \\ &= \operatorname{vol}(B(0,1)) \cdot (1-t) \cdot \left(\frac{1-t}{\sqrt{1-2t}}\right)^{n-1} \end{aligned}$$

$$\begin{aligned} \operatorname{vol}(\hat{E}') &= \operatorname{vol}(B(0,1)) \cdot |\det(\hat{L}')| \\ &= \operatorname{vol}(B(0,1)) \cdot ab^{n-1} \\ &= \operatorname{vol}(B(0,1)) \cdot (1-t) \cdot \left(\frac{1-t}{\sqrt{1-2t}}\right)^{n-1} \\ &= \operatorname{vol}(B(0,1)) \cdot \frac{(1-t)^n}{(\sqrt{1-2t})^{n-1}} \end{aligned}$$

$$\begin{aligned} \operatorname{vol}(\hat{E}') &= \operatorname{vol}(B(0,1)) \cdot |\det(\hat{L}')| \\ &= \operatorname{vol}(B(0,1)) \cdot ab^{n-1} \\ &= \operatorname{vol}(B(0,1)) \cdot (1-t) \cdot \left(\frac{1-t}{\sqrt{1-2t}}\right)^{n-1} \\ &= \operatorname{vol}(B(0,1)) \cdot \frac{(1-t)^n}{(\sqrt{1-2t})^{n-1}} \end{aligned}$$

We use the shortcut $\Phi := vol(B(0, 1))$.

 $\frac{\operatorname{d}\operatorname{vol}(\hat{E}')}{\operatorname{d}t}$

$$\frac{\operatorname{d} \operatorname{vol}(\hat{E}')}{\operatorname{d} t} = \frac{\operatorname{d}}{\operatorname{d} t} \left(\Phi \frac{(1-t)^n}{(\sqrt{1-2t})^{n-1}} \right)$$

$$\frac{\operatorname{d}\operatorname{vol}(\hat{E}')}{\operatorname{d}t} = \frac{\operatorname{d}}{\operatorname{d}t} \left(\Phi \frac{(1-t)^n}{(\sqrt{1-2t})^{n-1}} \right)$$
$$= \frac{\Phi}{N^2}$$

$$N = \operatorname{denominator}$$

$$\frac{\mathrm{d}\operatorname{vol}(\hat{E}')}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t} \left(\Phi \frac{(1-t)^n}{(\sqrt{1-2t})^{n-1}} \right)$$
$$= \frac{\Phi}{N^2} \cdot \left(\frac{(-1) \cdot n(1-t)^{n-1}}{\text{derivative of numerator}} \right)$$

$$\begin{split} \frac{\mathrm{d}\operatorname{vol}(\hat{E}')}{\mathrm{d}\,t} &= \frac{\mathrm{d}}{\mathrm{d}\,t} \left(\Phi \frac{(1-t)^n}{(\sqrt{1-2t})^{n-1}} \right) \\ &= \frac{\Phi}{N^2} \cdot \left((-1) \cdot n (1-t)^{n-1} \cdot (\sqrt{1-2t})^{n-1} \right. \\ &\qquad \qquad \left. \left(\mathrm{denominator} \right) \right] \end{split}$$

$$\begin{split} \frac{\mathrm{d}\operatorname{vol}(\hat{E}')}{\mathrm{d}\,t} &= \frac{\mathrm{d}}{\mathrm{d}\,t} \left(\Phi \frac{(1-t)^n}{(\sqrt{1-2t})^{n-1}} \right) \\ &= \frac{\Phi}{N^2} \cdot \left((-1) \cdot n(1-t)^{n-1} \cdot (\sqrt{1-2t})^{n-1} \right. \\ &\left. - (n-1)(\sqrt{1-2t})^{n-2} \cdot \frac{1}{2\sqrt{1-2t}} \cdot (-2) \cdot \frac{(1-t)^n}{\text{numerator}} \right] \end{split}$$

$$\begin{split} \frac{\mathrm{d} \operatorname{vol}(\hat{E}')}{\mathrm{d} \, t} &= \frac{\mathrm{d}}{\mathrm{d} \, t} \left(\Phi \frac{(1-t)^n}{(\sqrt{1-2t})^{n-1}} \right) \\ &= \frac{\Phi}{N^2} \cdot \left((-1) \cdot n (1-t)^{n-1} \cdot (\sqrt{1-2t})^{n-1} \right. \\ &\left. - (n-1) (\sqrt{1-2t})^{n-2} \cdot \frac{1}{2\sqrt{1-2t}} \cdot (-2) \cdot (1-t)^n \right) \\ &= \frac{\Phi}{N^2} \cdot (\sqrt{1-2t})^{n-3} \cdot (1-t)^{n-1} \end{split}$$

$$\begin{split} \frac{\mathrm{d} \operatorname{vol}(\hat{E}')}{\mathrm{d} \, t} &= \frac{\mathrm{d}}{\mathrm{d} \, t} \left(\Phi \frac{(1-t)^n}{(\sqrt{1-2t})^{n-1}} \right) \\ &= \frac{\Phi}{N^2} \cdot \left((-1) \cdot n (1-t)^{n-1} \cdot (\sqrt{1-2t})^{n-1} \right. \\ &\left. - (n-1) (\sqrt{1-2t})^{n-2} \cdot \frac{1}{2\sqrt{1-2t}} \cdot (-2) \cdot (1-t)^n \right) \\ &= \frac{\Phi}{N^2} \cdot (\sqrt{1-2t})^{n-3} \cdot (1-t)^{n-1} \end{split}$$

$$\begin{split} \frac{\mathrm{d} \operatorname{vol}(\hat{E}')}{\mathrm{d} t} &= \frac{\mathrm{d}}{\mathrm{d} t} \left(\Phi \frac{(1-t)^n}{(\sqrt{1-2t})^{n-1}} \right) \\ &= \frac{\Phi}{N^2} \cdot \left((-1) \cdot n(1-t)^{n-1} \cdot (\sqrt{1-2t})^{n-1} \right) \\ &- (n-1)(\sqrt{1-2t})^{n-2} \cdot \frac{1}{2\sqrt{1-2t}} \cdot (-2) \cdot (1-t)^n \right) \\ &= \frac{\Phi}{N^2} \cdot (\sqrt{1-2t})^{n-3} \cdot (1-t)^{n-1} \end{split}$$

$$\begin{split} \frac{\mathrm{d} \operatorname{vol}(\hat{E}')}{\mathrm{d} t} &= \frac{\mathrm{d}}{\mathrm{d} t} \left(\Phi \frac{(1-t)^n}{(\sqrt{1-2t})^{n-1}} \right) \\ &= \frac{\Phi}{N^2} \cdot \left((-1) \cdot n(1-t)^{n-1} \cdot (\sqrt{1-2t})^{n-1} \right) \\ &- (n-1)(\sqrt{1-2t})^{n-2} \cdot \frac{1}{2\sqrt{1-2t}} \cdot (-2) \cdot (1-t)^n \right) \\ &= \frac{\Phi}{N^2} \cdot (\sqrt{1-2t})^{n-3} \cdot (1-t)^{n-1} \end{split}$$

$$\begin{split} \frac{\mathrm{d} \operatorname{vol}(\hat{E}')}{\mathrm{d} t} &= \frac{\mathrm{d}}{\mathrm{d} t} \left(\Phi \frac{(1-t)^n}{(\sqrt{1-2t})^{n-1}} \right) \\ &= \frac{\Phi}{N^2} \cdot \left((-1) \cdot n (1-t)^{n-1} \cdot (\sqrt{1-2t})^{n-1} \right) \\ &- (n-1) (\sqrt{1-2t})^{n-2} \cdot \frac{1}{2\sqrt{1-2t}} \cdot (-2) \cdot (1-t)^n \right) \\ &= \frac{\Phi}{N^2} \cdot (\sqrt{1-2t})^{n-3} \cdot (1-t)^{n-1} \end{split}$$

$$\begin{split} \frac{\mathrm{d} \operatorname{vol}(\hat{E}')}{\mathrm{d} t} &= \frac{\mathrm{d}}{\mathrm{d} t} \left(\Phi \frac{(1-t)^n}{(\sqrt{1-2t})^{n-1}} \right) \\ &= \frac{\Phi}{N^2} \cdot \left((-1) \cdot n(1-t)^{n-1} \cdot (\sqrt{1-2t})^{n-1} \right) \\ &= (n-1)(\sqrt{1-2t})^{n-2} \cdot \frac{1}{Z\sqrt{1-2t}} \cdot (2) \cdot (1-t)^n \right) \\ &= \frac{\Phi}{N^2} \cdot (\sqrt{1-2t})^{n-3} \cdot (1-t)^{n-1} \end{split}$$

$$\frac{\mathrm{d}\operatorname{vol}(\hat{E}')}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t} \left(\Phi \frac{(1-t)^n}{(\sqrt{1-2t})^{n-1}} \right) \\
= \frac{\Phi}{N^2} \cdot \left((-1) \cdot n(1-t)^{n-1} \cdot (\sqrt{1-2t})^{n-1} \right) \\
= (n-1)(\sqrt{1-2t})^{n-2} \cdot \frac{1}{2\sqrt{1-2t}} \cdot (2) \cdot (1-t)^n \right) \\
= \frac{\Phi}{N^2} \cdot (\sqrt{1-2t})^{n-3} \cdot (1-t)^{n-1} \\
\cdot \left((n-1)(1-t) - n(1-2t) \right)$$

$$\frac{\mathrm{d}\operatorname{vol}(\hat{E}')}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t} \left(\Phi \frac{(1-t)^n}{(\sqrt{1-2t})^{n-1}} \right)$$

$$= \frac{\Phi}{N^2} \cdot \left((-1) \cdot n(1-t)^{n-1} \cdot (\sqrt{1-2t})^{n-1} \right)$$

$$= (n-1)(\sqrt{1-2t})^{n-2} \cdot \frac{1}{2\sqrt{1-2t}} \cdot (2) \cdot (1-t)^n \right)$$

$$= \frac{\Phi}{N^2} \cdot (\sqrt{1-2t})^{n-3} \cdot (1-t)^{n-1}$$

$$\cdot \left((n-1)(1-t) - n(1-2t) \right)$$

$$= \frac{\Phi}{N^2} \cdot (\sqrt{1-2t})^{n-3} \cdot (1-t)^{n-1} \cdot \left((n+1)t - 1 \right)$$

- We obtain the minimum for $t = \frac{1}{n+1}$.
- For this value we obtain

a

- We obtain the minimum for $t = \frac{1}{n+1}$.
- For this value we obtain

$$a = 1 - t$$

- We obtain the minimum for $t = \frac{1}{n+1}$.
- For this value we obtain

$$a = 1 - t = \frac{n}{n+1}$$

- We obtain the minimum for $t = \frac{1}{n+1}$.
- For this value we obtain

$$a=1-t=\frac{n}{n+1}$$
 and $b=$

- We obtain the minimum for $t = \frac{1}{n+1}$.
- For this value we obtain

$$a = 1 - t = \frac{n}{n+1}$$
 and $b = \frac{1-t}{\sqrt{1-2t}}$

- We obtain the minimum for $t = \frac{1}{n+1}$.
- For this value we obtain

$$a = 1 - t = \frac{n}{n+1}$$
 and $b = \frac{1-t}{\sqrt{1-2t}} = \frac{n}{\sqrt{n^2-1}}$

- We obtain the minimum for $t = \frac{1}{n+1}$.
- For this value we obtain

$$a = 1 - t = \frac{n}{n+1}$$
 and $b = \frac{1-t}{\sqrt{1-2t}} = \frac{n}{\sqrt{n^2-1}}$

To see the equation for b, observe that

 b^2

- We obtain the minimum for $t = \frac{1}{n+1}$.
- For this value we obtain

$$a = 1 - t = \frac{n}{n+1}$$
 and $b = \frac{1-t}{\sqrt{1-2t}} = \frac{n}{\sqrt{n^2-1}}$

$$b^2 = \frac{(1-t)^2}{1-2t}$$

- We obtain the minimum for $t = \frac{1}{n+1}$.
- For this value we obtain

$$a = 1 - t = \frac{n}{n+1}$$
 and $b = \frac{1-t}{\sqrt{1-2t}} = \frac{n}{\sqrt{n^2-1}}$

$$b^{2} = \frac{(1-t)^{2}}{1-2t} = \frac{(1-\frac{1}{n+1})^{2}}{1-\frac{2}{n+1}}$$

- We obtain the minimum for $t = \frac{1}{n+1}$.
- For this value we obtain

$$a = 1 - t = \frac{n}{n+1}$$
 and $b = \frac{1-t}{\sqrt{1-2t}} = \frac{n}{\sqrt{n^2-1}}$

$$b^{2} = \frac{(1-t)^{2}}{1-2t} = \frac{(1-\frac{1}{n+1})^{2}}{1-\frac{2}{n+1}} = \frac{(\frac{n}{n+1})^{2}}{\frac{n-1}{n+1}}$$

- We obtain the minimum for $t = \frac{1}{n+1}$.
- For this value we obtain

$$a = 1 - t = \frac{n}{n+1}$$
 and $b = \frac{1-t}{\sqrt{1-2t}} = \frac{n}{\sqrt{n^2-1}}$

$$b^{2} = \frac{(1-t)^{2}}{1-2t} = \frac{(1-\frac{1}{n+1})^{2}}{1-\frac{2}{n+1}} = \frac{(\frac{n}{n+1})^{2}}{\frac{n-1}{n+1}} = \frac{n^{2}}{n^{2}-1}$$

Let $\gamma_n=\frac{{\rm vol}(\hat E')}{{\rm vol}(B(0,1))}=ab^{n-1}$ be the ratio by which the volume changes:

$$\gamma_n^2$$

Let $y_n = \frac{\operatorname{vol}(\hat{E}')}{\operatorname{vol}(B(0,1))} = ab^{n-1}$ be the ratio by which the volume changes:

$$y_n^2 = \left(\frac{n}{n+1}\right)^2 \left(\frac{n^2}{n^2-1}\right)^{n-1}$$

Let $y_n = \frac{\operatorname{vol}(\vec{E}')}{\operatorname{vol}(B(0,1))} = ab^{n-1}$ be the ratio by which the volume changes:

$$y_n^2 = \left(\frac{n}{n+1}\right)^2 \left(\frac{n^2}{n^2 - 1}\right)^{n-1}$$
$$= \left(1 - \frac{1}{n+1}\right)^2 \left(1 + \frac{1}{(n-1)(n+1)}\right)^{n-1}$$

Let $y_n = \frac{\operatorname{vol}(\vec{E}')}{\operatorname{vol}(B(0,1))} = ab^{n-1}$ be the ratio by which the volume changes:

$$y_n^2 = \left(\frac{n}{n+1}\right)^2 \left(\frac{n^2}{n^2 - 1}\right)^{n-1}$$

$$= \left(1 - \frac{1}{n+1}\right)^2 \left(1 + \frac{1}{(n-1)(n+1)}\right)^{n-1}$$

$$\leq e^{-2\frac{1}{n+1}} \cdot e^{\frac{1}{n+1}}$$

Let $y_n = \frac{\operatorname{vol}(\vec{E}')}{\operatorname{vol}(B(0,1))} = ab^{n-1}$ be the ratio by which the volume changes:

$$y_n^2 = \left(\frac{n}{n+1}\right)^2 \left(\frac{n^2}{n^2 - 1}\right)^{n-1}$$

$$= \left(1 - \frac{1}{n+1}\right)^2 \left(1 + \frac{1}{(n-1)(n+1)}\right)^{n-1}$$

$$\leq e^{-2\frac{1}{n+1}} \cdot e^{\frac{1}{n+1}}$$

$$= e^{-\frac{1}{n+1}}$$

Let $y_n = \frac{\operatorname{vol}(\vec{E}')}{\operatorname{vol}(B(0,1))} = ab^{n-1}$ be the ratio by which the volume changes:

$$y_n^2 = \left(\frac{n}{n+1}\right)^2 \left(\frac{n^2}{n^2 - 1}\right)^{n-1}$$

$$= \left(1 - \frac{1}{n+1}\right)^2 \left(1 + \frac{1}{(n-1)(n+1)}\right)^{n-1}$$

$$\leq e^{-2\frac{1}{n+1}} \cdot e^{\frac{1}{n+1}}$$

$$= e^{-\frac{1}{n+1}}$$

where we used $(1+x)^a \le e^{ax}$ for $x \in \mathbb{R}$ and a > 0.

Let $y_n = \frac{\operatorname{vol}(\vec{E}')}{\operatorname{vol}(B(0,1))} = ab^{n-1}$ be the ratio by which the volume changes:

$$y_n^2 = \left(\frac{n}{n+1}\right)^2 \left(\frac{n^2}{n^2 - 1}\right)^{n-1}$$

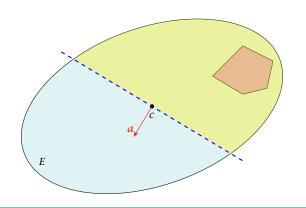
$$= \left(1 - \frac{1}{n+1}\right)^2 \left(1 + \frac{1}{(n-1)(n+1)}\right)^{n-1}$$

$$\le e^{-2\frac{1}{n+1}} \cdot e^{\frac{1}{n+1}}$$

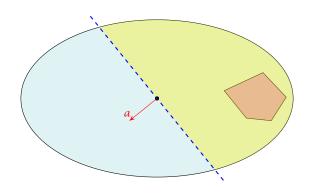
$$= e^{-\frac{1}{n+1}}$$

where we used $(1+x)^a \le e^{ax}$ for $x \in \mathbb{R}$ and a > 0.

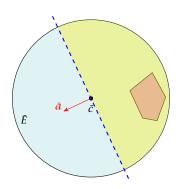
This gives $y_n \leq e^{-\frac{1}{2(n+1)}}$.



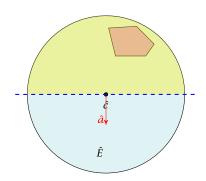
▶ Use f^{-1} (recall that f = Lx + t is the affine transformation of the unit ball) to translate/distort the ellipsoid (back) into the unit ball.



▶ Use f^{-1} (recall that f = Lx + t is the affine transformation of the unit ball) to translate/distort the ellipsoid (back) into the unit ball.

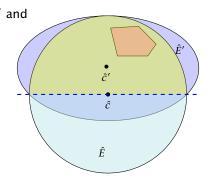


- ▶ Use f^{-1} (recall that f = Lx + t is the affine transformation of the unit ball) to translate/distort the ellipsoid (back) into the unit ball.
- Use a rotation R^{-1} to rotate the unit ball such that the normal vector of the halfspace is parallel to e_1 .

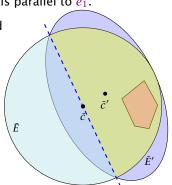


- ▶ Use f^{-1} (recall that f = Lx + t is the affine transformation of the unit ball) to translate/distort the ellipsoid (back) into the unit ball.
- Use a rotation R^{-1} to rotate the unit ball such that the normal vector of the halfspace is parallel to e_1 .

Compute the new center \hat{c}' and the new matrix \hat{Q}' for this simplified setting.



- ▶ Use f^{-1} (recall that f = Lx + t is the affine transformation of the unit ball) to translate/distort the ellipsoid (back) into the unit ball.
- Use a rotation R^{-1} to rotate the unit ball such that the normal vector of the halfspace is parallel to e_1 .
- ightharpoonup Compute the new center \hat{c}' and the new matrix \hat{O}' for this simplified setting.
- Use the transformations R and f to get the new center c' and the new matrix O' for the original ellipsoid *E*.

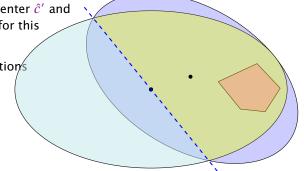


▶ Use f^{-1} (recall that f = Lx + t is the affine transformation of the unit ball) to translate/distort the ellipsoid (back) into the unit ball.

Use a rotation R^{-1} to rotate the unit ball such that the normal vector of the halfspace is parallel to e_1 .

Compute the new center \hat{c}' and the new matrix \hat{Q}' for this simplified setting.

P Use the transformations R and f to get the new center c' and the new matrix Q' for the original ellipsoid E.

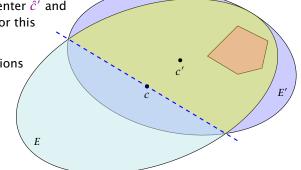


▶ Use f^{-1} (recall that f = Lx + t is the affine transformation of the unit ball) to translate/distort the ellipsoid (back) into the unit ball.

• Use a rotation R^{-1} to rotate the unit ball such that the normal vector of the halfspace is parallel to e_1 .

ightharpoonup Compute the new center \hat{c}' and the new matrix \hat{Q}' for this simplified setting.

Use the transformations R and f to get the new center c' and the new matrix O' for the original ellipsoid *E*.



$$e^{-\frac{1}{2(n+1)}}$$

$$e^{-\frac{1}{2(n+1)}} \ge \frac{\operatorname{vol}(\hat{E}')}{\operatorname{vol}(B(0,1))}$$

$$e^{-\frac{1}{2(n+1)}} \ge \frac{\operatorname{vol}(\hat{E}')}{\operatorname{vol}(B(0,1))} = \frac{\operatorname{vol}(\hat{E}')}{\operatorname{vol}(\hat{E})}$$

$$e^{-\frac{1}{2(n+1)}} \ge \frac{\operatorname{vol}(\hat{E}')}{\operatorname{vol}(B(0,1))} = \frac{\operatorname{vol}(\hat{E}')}{\operatorname{vol}(\hat{E})} = \frac{\operatorname{vol}(R(\hat{E}'))}{\operatorname{vol}(R(\hat{E}))}$$

$$\begin{split} e^{-\frac{1}{2(n+1)}} &\geq \frac{\operatorname{vol}(\hat{E}')}{\operatorname{vol}(B(0,1))} = \frac{\operatorname{vol}(\hat{E}')}{\operatorname{vol}(\hat{E})} = \frac{\operatorname{vol}(R(\hat{E}'))}{\operatorname{vol}(R(\hat{E}))} \\ &= \frac{\operatorname{vol}(\bar{E}')}{\operatorname{vol}(\bar{E})} \end{split}$$

$$e^{-\frac{1}{2(n+1)}} \ge \frac{\text{vol}(\hat{E}')}{\text{vol}(B(0,1))} = \frac{\text{vol}(\hat{E}')}{\text{vol}(\hat{E})} = \frac{\text{vol}(R(\hat{E}'))}{\text{vol}(R(\hat{E}))}$$
$$= \frac{\text{vol}(\bar{E}')}{\text{vol}(\bar{E})} = \frac{\text{vol}(f(\bar{E}'))}{\text{vol}(f(\bar{E}))}$$

$$e^{-\frac{1}{2(n+1)}} \ge \frac{\text{vol}(\hat{E}')}{\text{vol}(B(0,1))} = \frac{\text{vol}(\hat{E}')}{\text{vol}(\hat{E})} = \frac{\text{vol}(R(\hat{E}'))}{\text{vol}(R(\hat{E}))}$$
$$= \frac{\text{vol}(\bar{E}')}{\text{vol}(\bar{E})} = \frac{\text{vol}(f(\bar{E}'))}{\text{vol}(f(\bar{E}))} = \frac{\text{vol}(E')}{\text{vol}(E)}$$

$$e^{-\frac{1}{2(n+1)}} \ge \frac{\text{vol}(\hat{E}')}{\text{vol}(B(0,1))} = \frac{\text{vol}(\hat{E}')}{\text{vol}(\hat{E})} = \frac{\text{vol}(R(\hat{E}'))}{\text{vol}(R(\hat{E}))}$$
$$= \frac{\text{vol}(\bar{E}')}{\text{vol}(\bar{E})} = \frac{\text{vol}(f(\bar{E}'))}{\text{vol}(f(\bar{E}))} = \frac{\text{vol}(E')}{\text{vol}(E)}$$

Here it is important that mapping a set with affine function f(x) = Lx + t changes the volume by factor det(L).

How to compute the new parameters?

How to compute the new parameters?

The transformation function of the (old) ellipsoid: f(x) = Lx + c;

How to compute the new parameters?

The transformation function of the (old) ellipsoid: f(x) = Lx + c;

How to compute the new parameters?

The transformation function of the (old) ellipsoid: f(x) = Lx + c;

$$f^{-1}(H) = \{ f^{-1}(x) \mid a^{T}(x - c) \le 0 \}$$

How to compute the new parameters?

The transformation function of the (old) ellipsoid: f(x) = Lx + c;

$$f^{-1}(H) = \{ f^{-1}(x) \mid a^{T}(x - c) \le 0 \}$$
$$= \{ f^{-1}(f(y)) \mid a^{T}(f(y) - c) \le 0 \}$$

How to compute the new parameters?

The transformation function of the (old) ellipsoid: f(x) = Lx + c;

$$f^{-1}(H) = \{ f^{-1}(x) \mid a^{T}(x - c) \le 0 \}$$

$$= \{ f^{-1}(f(y)) \mid a^{T}(f(y) - c) \le 0 \}$$

$$= \{ y \mid a^{T}(f(y) - c) \le 0 \}$$

How to compute the new parameters?

The transformation function of the (old) ellipsoid: f(x) = Lx + c;

$$f^{-1}(H) = \{ f^{-1}(x) \mid a^{T}(x - c) \le 0 \}$$

$$= \{ f^{-1}(f(y)) \mid a^{T}(f(y) - c) \le 0 \}$$

$$= \{ y \mid a^{T}(f(y) - c) \le 0 \}$$

$$= \{ y \mid a^{T}(Ly + c - c) \le 0 \}$$

How to compute the new parameters?

The transformation function of the (old) ellipsoid: f(x) = Lx + c;

$$f^{-1}(H) = \{ f^{-1}(x) \mid a^{T}(x - c) \le 0 \}$$

$$= \{ f^{-1}(f(y)) \mid a^{T}(f(y) - c) \le 0 \}$$

$$= \{ y \mid a^{T}(f(y) - c) \le 0 \}$$

$$= \{ y \mid a^{T}(Ly + c - c) \le 0 \}$$

$$= \{ y \mid (a^{T}L)y \le 0 \}$$

How to compute the new parameters?

The transformation function of the (old) ellipsoid: f(x) = Lx + c;

The halfspace to be intersected: $H = \{x \mid a^T(x - c) \le 0\}$;

$$f^{-1}(H) = \{ f^{-1}(x) \mid a^{T}(x - c) \le 0 \}$$

$$= \{ f^{-1}(f(y)) \mid a^{T}(f(y) - c) \le 0 \}$$

$$= \{ y \mid a^{T}(f(y) - c) \le 0 \}$$

$$= \{ y \mid a^{T}(Ly + c - c) \le 0 \}$$

$$= \{ y \mid (a^{T}L)y \le 0 \}$$

This means $\bar{a} = L^T a$.

After rotating back (applying R^{-1}) the normal vector of the halfspace points in negative x_1 -direction. Hence,

$$R^{-1}\left(\frac{L^T a}{\|L^T a\|}\right) = -e_1 \quad \Rightarrow \quad -\frac{L^T a}{\|L^T a\|} = R \cdot e_1$$

After rotating back (applying R^{-1}) the normal vector of the halfspace points in negative x_1 -direction. Hence,

$$R^{-1}\left(\frac{L^T a}{\|L^T a\|}\right) = -e_1 \quad \Rightarrow \quad -\frac{L^T a}{\|L^T a\|} = R \cdot e_1$$

Hence,

 \bar{c}'

After rotating back (applying R^{-1}) the normal vector of the halfspace points in negative x_1 -direction. Hence,

$$R^{-1}\left(\frac{L^{T}a}{\|L^{T}a\|}\right) = -e_{1} \quad \Rightarrow \quad -\frac{L^{T}a}{\|L^{T}a\|} = R \cdot e_{1}$$

$$\bar{c}' = R \cdot \hat{c}'$$

After rotating back (applying R^{-1}) the normal vector of the halfspace points in negative x_1 -direction. Hence,

$$R^{-1}\left(\frac{L^T a}{\|L^T a\|}\right) = -e_1 \quad \Rightarrow \quad -\frac{L^T a}{\|L^T a\|} = R \cdot e_1$$

$$\bar{c}' = R \cdot \hat{c}' = R \cdot \frac{1}{n+1} e_1$$

After rotating back (applying R^{-1}) the normal vector of the halfspace points in negative x_1 -direction. Hence,

$$R^{-1}\left(\frac{L^{T}a}{\|L^{T}a\|}\right) = -e_{1} \quad \Rightarrow \quad -\frac{L^{T}a}{\|L^{T}a\|} = R \cdot e_{1}$$

$$\bar{c}' = R \cdot \hat{c}' = R \cdot \frac{1}{n+1} e_1 = -\frac{1}{n+1} \frac{L^T a}{\|L^T a\|}$$

After rotating back (applying \mathbb{R}^{-1}) the normal vector of the halfspace points in negative x_1 -direction. Hence,

$$R^{-1}\left(\frac{L^T a}{\|L^T a\|}\right) = -e_1 \quad \Rightarrow \quad -\frac{L^T a}{\|L^T a\|} = R \cdot e_1$$

Hence,

$$\bar{c}' = R \cdot \hat{c}' = R \cdot \frac{1}{n+1} e_1 = -\frac{1}{n+1} \frac{L^T a}{\|L^T a\|}$$

c'

After rotating back (applying R^{-1}) the normal vector of the halfspace points in negative x_1 -direction. Hence,

$$R^{-1}\left(\frac{L^{T}a}{\|L^{T}a\|}\right) = -e_{1} \quad \Rightarrow \quad -\frac{L^{T}a}{\|L^{T}a\|} = R \cdot e_{1}$$

$$\bar{c}' = R \cdot \hat{c}' = R \cdot \frac{1}{n+1} e_1 = -\frac{1}{n+1} \frac{L^T a}{\|L^T a\|}$$

$$c' = f(\bar{c}')$$

After rotating back (applying R^{-1}) the normal vector of the halfspace points in negative x_1 -direction. Hence,

$$R^{-1}\left(\frac{L^{T}a}{\|L^{T}a\|}\right) = -e_{1} \quad \Rightarrow \quad -\frac{L^{T}a}{\|L^{T}a\|} = R \cdot e_{1}$$

$$\bar{c}' = R \cdot \hat{c}' = R \cdot \frac{1}{n+1} e_1 = -\frac{1}{n+1} \frac{L^T a}{\|L^T a\|}$$

$$c' = f(\bar{c}') = L \cdot \bar{c}' + c$$

After rotating back (applying R^{-1}) the normal vector of the halfspace points in negative x_1 -direction. Hence,

$$R^{-1}\left(\frac{L^T a}{\|L^T a\|}\right) = -e_1 \quad \Rightarrow \quad -\frac{L^T a}{\|L^T a\|} = R \cdot e_1$$

$$\bar{c}' = R \cdot \hat{c}' = R \cdot \frac{1}{n+1} e_1 = -\frac{1}{n+1} \frac{L^T a}{\|L^T a\|}$$

$$c' = f(\bar{c}') = L \cdot \bar{c}' + c$$
$$= -\frac{1}{n+1} L \frac{L^T a}{\|L^T a\|} + c$$

After rotating back (applying \mathbb{R}^{-1}) the normal vector of the halfspace points in negative x_1 -direction. Hence,

$$R^{-1}\left(\frac{L^T a}{\|L^T a\|}\right) = -e_1 \quad \Rightarrow \quad -\frac{L^T a}{\|L^T a\|} = R \cdot e_1$$

$$\bar{c}' = R \cdot \hat{c}' = R \cdot \frac{1}{n+1} e_1 = -\frac{1}{n+1} \frac{L^T a}{\|L^T a\|}$$

$$c' = f(\bar{c}') = L \cdot \bar{c}' + c$$

$$= -\frac{1}{n+1} L \frac{L^T a}{\|L^T a\|} + c$$

$$= c - \frac{1}{n+1} \frac{Qa}{\sqrt{a^T Qa}}$$

For computing the matrix Q' of the new ellipsoid we assume in the following that \hat{E}' , \bar{E}' and E' refer to the ellispoids centered in the origin.

$$\hat{Q}' = \begin{pmatrix} a^2 & 0 & \dots & 0 \\ 0 & b^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & b^2 \end{pmatrix}$$

This gives

$$\hat{Q}' = \frac{n^2}{n^2 - 1} \left(I - \frac{2}{n+1} e_1 e_1^T \right)$$

$$\hat{Q}' = \begin{pmatrix} a^2 & 0 & \dots & 0 \\ 0 & b^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & b^2 \end{pmatrix}$$

This gives

$$\hat{Q}' = \frac{n^2}{n^2 - 1} \left(I - \frac{2}{n+1} e_1 e_1^T \right)$$

$$\hat{Q}' = \begin{pmatrix} a^2 & 0 & \dots & 0 \\ 0 & b^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & b^2 \end{pmatrix}$$

This gives

$$\hat{Q}' = \frac{n^2}{n^2 - 1} \left(I - \frac{2}{n+1} e_1 e_1^T \right)$$

$$\hat{Q}' = \begin{pmatrix} a^2 & 0 & \dots & 0 \\ 0 & b^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & b^2 \end{pmatrix}$$

This gives

$$\hat{Q}' = \frac{n^2}{n^2 - 1} \left(I - \frac{2}{n+1} e_1 e_1^T \right)$$

$$b^{2} - b^{2} \frac{2}{n+1} = \frac{n^{2}}{n^{2} - 1} \frac{2n^{2}}{(n-1)(n+1)^{2}}$$
$$= \frac{n^{2}(n+1) - 2n^{2}}{(n-1)(n+1)^{2}} = \frac{n^{2}(n-1)}{(n-1)(n+1)^{2}} = a$$

$$\hat{Q}' = \begin{pmatrix} a^2 & 0 & \dots & 0 \\ 0 & b^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & b^2 \end{pmatrix}$$

This gives

$$\hat{Q}' = \frac{n^2}{n^2 - 1} \left(I - \frac{2}{n+1} e_1 e_1^T \right)$$

$$b^{2} - b^{2} \frac{2}{n+1} = \frac{n^{2}}{n^{2} - 1} - \frac{2n^{2}}{(n-1)(n+1)^{2}}$$
$$= \frac{n^{2}(n+1) - 2n^{2}}{(n-1)(n+1)^{2}} = \frac{n^{2}(n-1)}{(n-1)(n+1)^{2}} = a$$

$$\hat{Q}' = \begin{pmatrix} a^2 & 0 & \dots & 0 \\ 0 & b^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & b^2 \end{pmatrix}$$

This gives

$$\hat{Q}' = \frac{n^2}{n^2 - 1} \left(I - \frac{2}{n+1} e_1 e_1^T \right)$$

$$b^{2} - b^{2} \frac{2}{n+1} = \frac{n^{2}}{n^{2} - 1} - \frac{2n^{2}}{(n-1)(n+1)^{2}}$$
$$= \frac{n^{2}(n+1) - 2n^{2}}{(n-1)(n+1)^{2}} = \frac{n^{2}(n-1)}{(n-1)(n+1)^{2}} = a^{2}$$

$$\hat{Q}' = \begin{pmatrix} a^2 & 0 & \dots & 0 \\ 0 & b^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & b^2 \end{pmatrix}$$

This gives

$$\hat{Q}' = \frac{n^2}{n^2 - 1} \left(I - \frac{2}{n+1} e_1 e_1^T \right)$$

$$b^{2} - b^{2} \frac{2}{n+1} = \frac{n^{2}}{n^{2} - 1} - \frac{2n^{2}}{(n-1)(n+1)^{2}}$$
$$= \frac{n^{2}(n+1) - 2n^{2}}{(n-1)(n+1)^{2}} = \frac{n^{2}(n-1)}{(n-1)(n+1)^{2}} = a^{2}$$

$$\hat{Q}' = \begin{pmatrix} a^2 & 0 & \dots & 0 \\ 0 & b^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & b^2 \end{pmatrix}$$

This gives

$$\hat{Q}' = \frac{n^2}{n^2 - 1} \left(I - \frac{2}{n+1} e_1 e_1^T \right)$$

$$b^{2} - b^{2} \frac{2}{n+1} = \frac{n^{2}}{n^{2} - 1} - \frac{2n^{2}}{(n-1)(n+1)^{2}}$$
$$= \frac{n^{2}(n+1) - 2n^{2}}{(n-1)(n+1)^{2}} = \frac{n^{2}(n-1)}{(n-1)(n+1)^{2}} = a^{2}$$

Ē

$$\bar{E}' = R(\hat{E}')$$

$$\bar{E}' = R(\hat{E}')$$

$$= \{R(x) \mid x^T \hat{Q}'^{-1} x \le 1\}$$

$$\begin{split} \bar{E}' &= R(\hat{E}') \\ &= \{ R(x) \mid x^T \hat{Q}'^{-1} x \le 1 \} \\ &= \{ y \mid (R^{-1} y)^T \hat{Q}'^{-1} R^{-1} y \le 1 \} \end{split}$$

$$\begin{split} \bar{E}' &= R(\hat{E}') \\ &= \{ R(x) \mid x^T \hat{Q}'^{-1} x \le 1 \} \\ &= \{ y \mid (R^{-1}y)^T \hat{Q}'^{-1} R^{-1} y \le 1 \} \\ &= \{ y \mid y^T (R^T)^{-1} \hat{Q}'^{-1} R^{-1} y \le 1 \} \end{split}$$

$$\bar{E}' = R(\hat{E}')
= \{R(x) \mid x^T \hat{Q}'^{-1} x \le 1\}
= \{y \mid (R^{-1}y)^T \hat{Q}'^{-1} R^{-1} y \le 1\}
= \{y \mid y^T (R^T)^{-1} \hat{Q}'^{-1} R^{-1} y \le 1\}
= \{y \mid y^T (R\hat{Q}' R^T)^{-1} y \le 1\}
= \{y \mid y^T (R\hat{Q}' R^T)^{-1} y \le 1\}$$

Hence,

 \bar{Q}'

$$\bar{Q}' = R\hat{Q}'R^T$$

$$\begin{split} \bar{Q}' &= R\hat{Q}'R^T \\ &= R \cdot \frac{n^2}{n^2 - 1} \Big(I - \frac{2}{n+1} e_1 e_1^T \Big) \cdot R^T \end{split}$$

$$\begin{split} \bar{Q}' &= R\hat{Q}'R^T \\ &= R \cdot \frac{n^2}{n^2 - 1} \Big(I - \frac{2}{n+1} e_1 e_1^T \Big) \cdot R^T \\ &= \frac{n^2}{n^2 - 1} \Big(R \cdot R^T - \frac{2}{n+1} (Re_1) (Re_1)^T \Big) \end{split}$$

$$\begin{split} \bar{Q}' &= R \hat{Q}' R^T \\ &= R \cdot \frac{n^2}{n^2 - 1} \Big(I - \frac{2}{n+1} e_1 e_1^T \Big) \cdot R^T \\ &= \frac{n^2}{n^2 - 1} \Big(R \cdot R^T - \frac{2}{n+1} (Re_1) (Re_1)^T \Big) \\ &= \frac{n^2}{n^2 - 1} \Big(I - \frac{2}{n+1} \frac{L^T a a^T L}{\|L^T a\|^2} \Big) \end{split}$$

E'

$$E' = L(\bar{E}')$$

$$E' = L(\bar{E}')$$
= $\{L(x) \mid x^T \bar{Q}'^{-1} x \le 1\}$

$$E' = L(\bar{E}')$$

$$= \{L(x) \mid x^T \bar{Q}'^{-1} x \le 1\}$$

$$= \{y \mid (L^{-1}y)^T \bar{Q}'^{-1} L^{-1} y \le 1\}$$

$$E' = L(\bar{E}')$$

$$= \{L(x) \mid x^T \bar{Q}'^{-1} x \le 1\}$$

$$= \{y \mid (L^{-1}y)^T \bar{Q}'^{-1} L^{-1} y \le 1\}$$

$$= \{y \mid y^T (L^T)^{-1} \bar{Q}'^{-1} L^{-1} y \le 1\}$$

$$E' = L(\bar{E}')$$

$$= \{L(x) \mid x^T \bar{Q}'^{-1} x \le 1\}$$

$$= \{y \mid (L^{-1}y)^T \bar{Q}'^{-1} L^{-1} y \le 1\}$$

$$= \{y \mid y^T (L^T)^{-1} \bar{Q}'^{-1} L^{-1} y \le 1\}$$

$$= \{y \mid y^T (L\bar{Q}' L^T)^{-1} y \le 1\}$$

Hence,

Q'

Hence,

$$Q' = L\bar{Q}'L^T$$

Hence,

$$\begin{aligned} Q' &= L\bar{Q}'L^T \\ &= L \cdot \frac{n^2}{n^2 - 1} \Big(I - \frac{2}{n+1} \frac{L^T a a^T L}{a^T Q a} \Big) \cdot L^T \end{aligned}$$

Hence,

$$Q' = L\bar{Q}'L^{T}$$

$$= L \cdot \frac{n^{2}}{n^{2} - 1} \left(I - \frac{2}{n+1} \frac{L^{T}aa^{T}L}{a^{T}Qa} \right) \cdot L^{T}$$

$$= \frac{n^{2}}{n^{2} - 1} \left(Q - \frac{2}{n+1} \frac{Qaa^{T}Q}{a^{T}Qa} \right)$$

Incomplete Algorithm

Algorithm 1 ellipsoid-algorithm

- 1: **input:** point $c \in \mathbb{R}^n$, convex set $K \subseteq \mathbb{R}^n$
- 2: **output:** point $x \in K$ or "K is empty"
- 3: *Q* ← ???
- 4: repeat
- 5: if $c \in K$ then return c
- 6: **else**
- 7: choose a violated hyperplane *a*
- 8: $c \leftarrow c \frac{1}{n+1} \frac{Qa}{\sqrt{a^T Qa}}$
 - $Q \leftarrow \frac{n^2}{n^2 1} \left(Q \frac{2}{n+1} \frac{Qaa^T Q}{a^T Qa} \right)$
- 10: **endif**
- 11: until ???
- 12: return "K is empty"

Repeat: Size of basic solutions

Lemma 52

Let $P = \{x \in \mathbb{R}^n \mid Ax \leq b\}$ be a bounded polyhedron. Let $L := 2\langle A \rangle + \langle b \rangle + 2n(1 + \log_2 n)$. Then every entry x_j in a basic solution fulfills $|x_j| = \frac{D_j}{D}$ with $D_j, D \leq 2^L$.

In the following we use $\delta := 2^L$.

Proof:

We can replace P by $P':=\{x\mid A'x\leq b;x\geq 0\}$ where $A'=\begin{bmatrix}A-A\end{bmatrix}$. The lemma follows by applying Lemma 47, and observing that $\langle A'\rangle=2\langle A\rangle$ and n'=2n.

Repeat: Size of basic solutions

Lemma 52

Let $P = \{x \in \mathbb{R}^n \mid Ax \le b\}$ be a bounded polyhedron. Let $L := 2\langle A \rangle + \langle b \rangle + 2n(1 + \log_2 n)$. Then every entry x_j in a basic solution fulfills $|x_j| = \frac{D_j}{D}$ with $D_j, D \le 2^L$.

In the following we use $\delta := 2^L$.

Proof:

We can replace P by $P':=\{x\mid A'x\leq b; x\geq 0\}$ where $A'=\begin{bmatrix}A-A\end{bmatrix}$. The lemma follows by applying Lemma 47, and observing that $\langle A'\rangle=2\langle A\rangle$ and n'=2n.

For feasibility checking we can assume that the polytop P is bounded; it is sufficient to consider basic solutions.

Every entry x_i in a basic solution fulfills $|x_i| \le \delta$.

Hence, P is contained in the cube $-\delta \le x_i \le \delta$.

A vector in this cube has at most distance $R:=\sqrt{n}\delta$ from the origin.

Starting with the ball $E_0 := B(0,R)$ ensures that P is completely contained in the initial ellipsoid. This ellipsoid has volume at most $R^n \operatorname{vol}(B(0,1)) \le (n\delta)^n \operatorname{vol}(B(0,1))$.

For feasibility checking we can assume that the polytop P is bounded; it is sufficient to consider basic solutions.

Every entry x_i in a basic solution fulfills $|x_i| \le \delta$.

Hence, P is contained in the cube $-\delta \leq x_i \leq \delta$.

A vector in this cube has at most distance $R:=\sqrt{n}\delta$ from the origin.

Starting with the ball $E_0 := B(0,R)$ ensures that P is completely contained in the initial ellipsoid. This ellipsoid has volume at most $R^n \operatorname{vol}(B(0,1)) \le (n\delta)^n \operatorname{vol}(B(0,1))$.

For feasibility checking we can assume that the polytop P is bounded; it is sufficient to consider basic solutions.

Every entry x_i in a basic solution fulfills $|x_i| \le \delta$.

Hence, P is contained in the cube $-\delta \leq x_i \leq \delta$.

A vector in this cube has at most distance $R:=\sqrt{n}\delta$ from the origin.

Starting with the ball $E_0 := B(0,R)$ ensures that P is completely contained in the initial ellipsoid. This ellipsoid has volume at $most R^n vol(B(0,1)) \le (n\delta)^n vol(B(0,1))$.

For feasibility checking we can assume that the polytop P is bounded; it is sufficient to consider basic solutions.

Every entry x_i in a basic solution fulfills $|x_i| \le \delta$.

Hence, P is contained in the cube $-\delta \le x_i \le \delta$.

A vector in this cube has at most distance $R:=\sqrt{n}\delta$ from the origin.

Starting with the ball $E_0 := B(0,R)$ ensures that P is completely contained in the initial ellipsoid. This ellipsoid has volume at most $R^n \operatorname{vol}(B(0,1)) \le (n\delta)^n \operatorname{vol}(B(0,1))$.

For feasibility checking we can assume that the polytop P is bounded; it is sufficient to consider basic solutions.

Every entry x_i in a basic solution fulfills $|x_i| \le \delta$.

Hence, P is contained in the cube $-\delta \le x_i \le \delta$.

A vector in this cube has at most distance $R:=\sqrt{n}\delta$ from the origin.

Starting with the ball $E_0 := B(0,R)$ ensures that P is completely contained in the initial ellipsoid. This ellipsoid has volume at $most R^n vol(B(0,1)) \le (n\delta)^n vol(B(0,1))$.

For feasibility checking we can assume that the polytop P is bounded; it is sufficient to consider basic solutions.

Every entry x_i in a basic solution fulfills $|x_i| \le \delta$.

Hence, P is contained in the cube $-\delta \le x_i \le \delta$.

A vector in this cube has at most distance $R := \sqrt{n}\delta$ from the origin.

Starting with the ball $E_0 := B(0,R)$ ensures that P is completely contained in the initial ellipsoid. This ellipsoid has volume at most $R^n \operatorname{vol}(B(0,1)) \le (n\delta)^n \operatorname{vol}(B(0,1))$.

When can we terminate?

Let $P := \{x \mid Ax \leq b\}$ with $A \in \mathbb{Z}$ and $b \in \mathbb{Z}$ be a bounded polytop.

Consider the following polyhedron

$$P_{\lambda} := \left\{ x \mid Ax \le b + \frac{1}{\lambda} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \right\} ,$$

where $\lambda = \delta^2 + 1$.

Note that the volume of P_{λ} cannot be 0

When can we terminate?

Let $P := \{x \mid Ax \leq b\}$ with $A \in \mathbb{Z}$ and $b \in \mathbb{Z}$ be a bounded polytop.

Consider the following polyhedron

$$P_{\lambda} := \left\{ x \mid Ax \leq b + \frac{1}{\lambda} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \right\} ,$$

where $\lambda = \delta^2 + 1$.

Note that the volume of P_{λ} cannot be 0

When can we terminate?

Let $P := \{x \mid Ax \leq b\}$ with $A \in \mathbb{Z}$ and $b \in \mathbb{Z}$ be a bounded polytop.

Consider the following polyhedron

$$P_{\lambda} := \left\{ x \mid Ax \leq b + \frac{1}{\lambda} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \right\} ,$$

where $\lambda = \delta^2 + 1$.

Note that the volume of P_{λ} cannot be 0

Lemma 53

 P_{λ} is feasible if and only if P is feasible.

⇒: obvious

Lemma 53

 P_{λ} is feasible if and only if P is feasible.

←: obvious!

 \Longrightarrow

Consider the polyhedrons

$$\bar{P} = \left\{ x \mid \left[A - A \, I_m \right] x = b; x \ge 0 \right\}$$

and

$$\bar{P}_{\lambda} = \left\{ x \mid \left[A - A I_m \right] x = b + \frac{1}{\lambda} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}; x \ge 0 \right\} .$$

P is feasible if and only if \bar{P} is feasible, and P_{λ} feasible if and only if \bar{P}_{λ} feasible.

 \bar{P}_{λ} is bounded since P_{λ} and P are bounded

⇒:

Consider the polyhedrons

$$\bar{P} = \left\{ x \mid \left[A - A \, I_m \right] x = b; x \ge 0 \right\}$$

and

$$\bar{P}_{\lambda} = \left\{ x \mid \left[A - A I_m \right] x = b + \frac{1}{\lambda} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}; x \ge 0 \right\}.$$

P is feasible if and only if \bar{P} is feasible, and P_{λ} feasible if and only if \bar{P}_{λ} feasible.

 $ar{P}_\lambda$ is bounded since P_λ and P are bounded

⇒:

Consider the polyhedrons

$$\bar{P} = \left\{ x \mid \left[A - A \, I_m \right] x = b; x \ge 0 \right\}$$

and

$$\bar{P}_{\lambda} = \left\{ x \mid \left[A - A I_m \right] x = b + \frac{1}{\lambda} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}; x \ge 0 \right\}.$$

P is feasible if and only if \bar{P} is feasible, and P_{λ} feasible if and only if \bar{P}_{λ} feasible.

 \bar{P}_{λ} is bounded since P_{λ} and P are bounded

⇒:

Consider the polyhedrons

$$\bar{P} = \left\{ x \mid \left[A - A I_m \right] x = b; x \ge 0 \right\}$$

and

$$\bar{P}_{\lambda} = \left\{ x \mid \left[A - A I_m \right] x = b + \frac{1}{\lambda} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}; x \ge 0 \right\}.$$

P is feasible if and only if \bar{P} is feasible, and P_{λ} feasible if and only if \bar{P}_{λ} feasible.

 \bar{P}_{λ} is bounded since P_{λ} and P are bounded.

Let
$$\bar{A} = [A - A I_m]$$
.

 $ar{P}_{\lambda}$ feasible implies that there is a basic feasible solution represented by

$$x_B = \bar{A}_B^{-1}b + \frac{1}{\lambda}\bar{A}_B^{-1}\begin{pmatrix} 1\\ \vdots\\ 1\end{pmatrix}$$

(The other *x*-values are zero)

The only reason that this basic feasible solution is not feasible for \bar{P} is that one of the basic variables becomes negative.

Hence, there exists i with

$$(\bar{A}_B^{-1}b)_i < 0 \leq (\bar{A}_B^{-1}b)_i + \frac{1}{\lambda}(\bar{A}_B^{-1}\vec{1})_i$$

Let
$$\bar{A} = [A - A I_m]$$
.

 \bar{P}_{λ} feasible implies that there is a basic feasible solution represented by

$$x_B = \bar{A}_B^{-1}b + \frac{1}{\lambda}\bar{A}_B^{-1}\begin{pmatrix} 1\\ \vdots\\ 1\end{pmatrix}$$

(The other x-values are zero)

The only reason that this basic feasible solution is not feasible for \bar{P} is that one of the basic variables becomes negative.

Hence, there exists i with

$$(\bar{A}_B^{-1}b)_i < 0 \leq (\bar{A}_B^{-1}b)_i + \frac{1}{\lambda}(\bar{A}_B^{-1}\vec{1})_i$$

Let
$$\bar{A} = [A - A I_m]$$
.

 $ar{P}_{\lambda}$ feasible implies that there is a basic feasible solution represented by

$$x_B = \bar{A}_B^{-1}b + \frac{1}{\lambda}\bar{A}_B^{-1}\begin{pmatrix} 1\\ \vdots\\ 1\end{pmatrix}$$

(The other x-values are zero)

The only reason that this basic feasible solution is not feasible for \bar{P} is that one of the basic variables becomes negative.

Hence, there exists i with

$$(\bar{A}_B^{-1}b)_i < 0 \le (\bar{A}_B^{-1}b)_i + \frac{1}{\lambda}(\bar{A}_B^{-1}\vec{1})_i$$

By Cramers rule we get

$$(\bar{A}_B^{-1}b)_i < 0 \quad \Longrightarrow \quad (\bar{A}_B^{-1}b)_i \le -\frac{1}{\det(\bar{A}_B)} \le -1/\delta$$

and

$$(\bar{A}_B^{-1}\vec{1})_i \leq \det(\bar{A}_B^j) \leq \delta$$
 ,

where \bar{A}_B^j is obtained by replacing the j-th column of \bar{A}_B by $\vec{1}$.

But then

$$(\bar{A}_B^{-1}b)_i + \frac{1}{\lambda}(\bar{A}_B^{-1}\vec{1})_i \leq -1/\delta + \delta/\lambda < 0$$

as we chose $\lambda = \delta^2 + 1$. Contradiction.



By Cramers rule we get

$$(\bar{A}_B^{-1}b)_i < 0 \implies (\bar{A}_B^{-1}b)_i \le -\frac{1}{\det(\bar{A}_B)} \le -1/\delta$$

and

$$(\bar{A}_B^{-1}\vec{1})_i \leq \det(\bar{A}_B^j) \leq \delta$$
 ,

where \bar{A}_B^j is obtained by replacing the j-th column of \bar{A}_B by $\vec{1}$.

But then

$$(\bar{A}_B^{-1}b)_i + rac{1}{\lambda}(\bar{A}_B^{-1}\vec{1})_i \leq -1/\delta + \delta/\lambda < 0$$
 ,

as we chose $\lambda = \delta^2 + 1$. Contradiction.

If P_{λ} is feasible then it contains a ball of radius $r:=1/\delta^3$. This has a volume of at least $r^n \mathrm{vol}(B(0,1)) = \frac{1}{\delta^{3n}} \mathrm{vol}(B(0,1))$.

If P_{λ} is feasible then it contains a ball of radius $r:=1/\delta^3$. This has a volume of at least $r^n \mathrm{vol}(B(0,1)) = \frac{1}{\delta^{3n}} \mathrm{vol}(B(0,1))$.

Proof:

If P_{λ} feasible then also P. Let x be feasible for P.

If P_{λ} is feasible then it contains a ball of radius $r:=1/\delta^3$. This has a volume of at least $r^n \mathrm{vol}(B(0,1)) = \frac{1}{\delta^{3n}} \mathrm{vol}(B(0,1))$.

Proof:

If P_{λ} feasible then also P. Let x be feasible for P.

This means $Ax \leq b$.

If P_{λ} is feasible then it contains a ball of radius $r:=1/\delta^3$. This has a volume of at least $r^n \mathrm{vol}(B(0,1)) = \frac{1}{\delta^{3n}} \mathrm{vol}(B(0,1))$.

Proof:

If P_{λ} feasible then also P. Let x be feasible for P. This means $Ax \leq b$.

$$(A(x+\vec{\ell}))_i$$

If P_{λ} is feasible then it contains a ball of radius $r:=1/\delta^3$. This has a volume of at least $r^n \mathrm{vol}(B(0,1)) = \frac{1}{\delta^{3n}} \mathrm{vol}(B(0,1))$.

Proof:

If P_{λ} feasible then also P. Let x be feasible for P. This means $Ax \leq b$.

$$(A(x+\vec{\ell}))_i = (Ax)_i + (A\vec{\ell})_i$$

If P_{λ} is feasible then it contains a ball of radius $r:=1/\delta^3$. This has a volume of at least $r^n \mathrm{vol}(B(0,1)) = \frac{1}{\delta^{3n}} \mathrm{vol}(B(0,1))$.

Proof:

If P_{λ} feasible then also P. Let x be feasible for P. This means $Ax \leq b$.

$$(A(x+\vec{\ell}))_i = (Ax)_i + (A\vec{\ell})_i \le b_i + \vec{a}_i^T \vec{\ell}$$

If P_{λ} is feasible then it contains a ball of radius $r:=1/\delta^3$. This has a volume of at least $r^n \mathrm{vol}(B(0,1)) = \frac{1}{\delta^{3n}} \mathrm{vol}(B(0,1))$.

Proof:

If P_{λ} feasible then also P. Let x be feasible for P. This means $Ax \leq b$.

$$(A(x + \vec{\ell}))_i = (Ax)_i + (A\vec{\ell})_i \le b_i + \vec{a}_i^T \vec{\ell}$$

$$\le b_i + ||\vec{a}_i|| \cdot ||\vec{\ell}||$$

If P_{λ} is feasible then it contains a ball of radius $r:=1/\delta^3$. This has a volume of at least $r^n \mathrm{vol}(B(0,1)) = \frac{1}{\delta^{3n}} \mathrm{vol}(B(0,1))$.

Proof:

If P_{λ} feasible then also P. Let x be feasible for P. This means $Ax \leq b$.

Let $\vec{\ell}$ with $\|\vec{\ell}\| \leq r$. Then

$$\begin{split} (A(x+\vec{\ell}))_i &= (Ax)_i + (A\vec{\ell})_i \leq b_i + \vec{a}_i^T \vec{\ell} \\ &\leq b_i + \|\vec{a}_i\| \cdot \|\vec{\ell}\| \leq b_i + \sqrt{n} \cdot 2^{\langle a_{\text{max}} \rangle} \cdot r \end{split}$$

If P_{λ} is feasible then it contains a ball of radius $r:=1/\delta^3$. This has a volume of at least $r^n \mathrm{vol}(B(0,1)) = \frac{1}{\delta^{3n}} \mathrm{vol}(B(0,1))$.

Proof:

If P_{λ} feasible then also P. Let x be feasible for P. This means $Ax \leq b$.

Let $\vec{\ell}$ with $\|\vec{\ell}\| \leq r$. Then

$$\begin{split} (A(x+\vec{\ell}))_i &= (Ax)_i + (A\vec{\ell})_i \le b_i + \vec{a}_i^T \vec{\ell} \\ &\le b_i + \|\vec{a}_i\| \cdot \|\vec{\ell}\| \le b_i + \sqrt{n} \cdot 2^{\langle a_{\max} \rangle} \cdot r \\ &\le b_i + \frac{\sqrt{n} \cdot 2^{\langle a_{\max} \rangle}}{\delta^3} \end{split}$$

If P_{λ} is feasible then it contains a ball of radius $r:=1/\delta^3$. This has a volume of at least $r^n \mathrm{vol}(B(0,1)) = \frac{1}{\delta^{3n}} \mathrm{vol}(B(0,1))$.

Proof:

If P_{λ} feasible then also P. Let x be feasible for P. This means $Ax \leq b$.

Let $\vec{\ell}$ with $\|\vec{\ell}\| \leq r$. Then

$$\begin{split} (A(x+\vec{\ell}))_i &= (Ax)_i + (A\vec{\ell})_i \le b_i + \vec{a}_i^T \vec{\ell} \\ &\le b_i + \|\vec{a}_i\| \cdot \|\vec{\ell}\| \le b_i + \sqrt{n} \cdot 2^{\langle a_{\text{max}} \rangle} \cdot r \\ &\le b_i + \frac{\sqrt{n} \cdot 2^{\langle a_{\text{max}} \rangle}}{\delta^3} \le b_i + \frac{1}{\delta^2 + 1} \le b_i + \frac{1}{\lambda} \end{split}$$

If P_{λ} is feasible then it contains a ball of radius $r:=1/\delta^3$. This has a volume of at least $r^n \mathrm{vol}(B(0,1)) = \frac{1}{\delta^{3n}} \mathrm{vol}(B(0,1))$.

Proof:

If P_{λ} feasible then also P. Let x be feasible for P. This means $Ax \leq b$.

Let $\vec{\ell}$ with $\|\vec{\ell}\| \leq r$. Then

$$\begin{split} (A(x+\vec{\ell}))_i &= (Ax)_i + (A\vec{\ell})_i \le b_i + \vec{a}_i^T \vec{\ell} \\ &\le b_i + \|\vec{a}_i\| \cdot \|\vec{\ell}\| \le b_i + \sqrt{n} \cdot 2^{\langle a_{\max} \rangle} \cdot r \\ &\le b_i + \frac{\sqrt{n} \cdot 2^{\langle a_{\max} \rangle}}{\delta^3} \le b_i + \frac{1}{\delta^2 + 1} \le b_i + \frac{1}{\lambda} \end{split}$$

Hence, $x + \vec{\ell}$ is feasible for P_{λ} which proves the lemma.

$$e^{-\frac{i}{2(n+1)}} \cdot \operatorname{vol}(B(0,R)) < \operatorname{vol}(B(0,r))$$

$$e^{-\frac{i}{2(n+1)}} \cdot \operatorname{vol}(B(0,R)) < \operatorname{vol}(B(0,r))$$

Hence,

i

$$e^{-\frac{i}{2(n+1)}}\cdot \operatorname{vol}(B(0,R)) < \operatorname{vol}(B(0,r))$$

$$i > 2(n+1)\ln\left(\frac{\operatorname{vol}(B(0,R))}{\operatorname{vol}(B(0,r))}\right)$$

$$e^{-\frac{i}{2(n+1)}}\cdot \operatorname{vol}(B(0,R)) < \operatorname{vol}(B(0,r))$$

$$i > 2(n+1)\ln\left(\frac{\operatorname{vol}(B(0,R))}{\operatorname{vol}(B(0,r))}\right)$$
$$= 2(n+1)\ln\left(n^n\delta^n \cdot \delta^{3n}\right)$$

$$e^{-\frac{i}{2(n+1)}}\cdot \operatorname{vol}(B(0,R)) < \operatorname{vol}(B(0,r))$$

$$i > 2(n+1)\ln\left(\frac{\operatorname{vol}(B(0,R))}{\operatorname{vol}(B(0,r))}\right)$$
$$= 2(n+1)\ln\left(n^n\delta^n \cdot \delta^{3n}\right)$$
$$= 8n(n+1)\ln(\delta) + 2(n+1)n\ln(n)$$

$$e^{-\frac{i}{2(n+1)}}\cdot \operatorname{vol}(B(0,R)) < \operatorname{vol}(B(0,r))$$

$$i > 2(n+1)\ln\left(\frac{\operatorname{vol}(B(0,R))}{\operatorname{vol}(B(0,r))}\right)$$

$$= 2(n+1)\ln\left(n^n\delta^n \cdot \delta^{3n}\right)$$

$$= 8n(n+1)\ln(\delta) + 2(n+1)n\ln(n)$$

$$= \mathcal{O}(\operatorname{poly}(n) \cdot L)$$

Algorithm 1 ellipsoid-algorithm

1: **input:** point $c \in \mathbb{R}^n$, convex set $K \subseteq \mathbb{R}^n$, radii R and r

2: with
$$K \subseteq B(c,R)$$
, and $B(x,r) \subseteq K$ for some x
3: **output:** point $x \in K$ or " K is empty"

4:
$$Q \leftarrow \operatorname{diag}(R^2, \dots, R^2) // \text{ i.e., } L = \operatorname{diag}(R, \dots, R)$$

5: **repeat**
6: **if**
$$c \in K$$
 then return c

endif

13: return "K is empty"

choose a violated hyperplane
$$a$$

choose a violated hy
$$1 \quad Qa$$

$$c \leftarrow c - \frac{1}{m+1} \frac{Qa}{\sqrt{\pi}}$$

$$c \leftarrow c - \frac{1}{n+1} \frac{Qa}{\sqrt{a^T Qa}}$$

$$c \leftarrow c - \frac{1}{n+1} \frac{Qa}{\sqrt{a^T Qa}}$$

$$n + 1 \sqrt{a^T Q a}$$

$$Q \leftarrow \frac{n^2}{n^2 - 1} \left(Q - \frac{2}{n+1} \frac{Q a a^T Q}{a^T Q a} \right)$$

$$n+1$$

12: **until** $\det(Q) \leq r^{2n}$ // i.e., $\det(L) \leq r^n$

$$+ 1 \sqrt{a^T}$$

Let $K \subseteq \mathbb{R}^n$ be a convex set. A separation oracle for K is an algorithm A that gets as input a point $x \in \mathbb{R}^n$ and either

- ightharpoonup certifies that $x \in K$,
- ightharpoonup or finds a hyperplane separating x from K.

We will usually assume that A is a polynomial-time algorithm.

In order to find a point in K we need

- a guarantee that a ball of radius x is contained inned in
- an initial ball 31 c. 31 with radius 4 that contains
- a separation oracle for

Let $K \subseteq \mathbb{R}^n$ be a convex set. A separation oracle for K is an algorithm A that gets as input a point $x \in \mathbb{R}^n$ and either

- ightharpoonup certifies that $x \in K$,
- or finds a hyperplane separating x from K.

We will usually assume that A is a polynomial-time algorithm.

In order to find a point in K we need

- a guarantee that a ball of radius of semialised that
- an hittal bail with ratifus a that contains
- a separation oracle for

Let $K \subseteq \mathbb{R}^n$ be a convex set. A separation oracle for K is an algorithm A that gets as input a point $x \in \mathbb{R}^n$ and either

- ightharpoonup certifies that $x \in K$,
- or finds a hyperplane separating x from K.

We will usually assume that A is a polynomial-time algorithm.

In order to find a point in K we need

Let $K \subseteq \mathbb{R}^n$ be a convex set. A separation oracle for K is an algorithm A that gets as input a point $x \in \mathbb{R}^n$ and either

- ightharpoonup certifies that $x \in K$,
- or finds a hyperplane separating x from K.

We will usually assume that A is a polynomial-time algorithm.

In order to find a point in K we need

- ightharpoonup a guarantee that a ball of radius r is contained in K,
- ightharpoonup an initial ball B(c,R) with radius R that contains K,
- ightharpoonup a separation oracle for K.

Let $K \subseteq \mathbb{R}^n$ be a convex set. A separation oracle for K is an algorithm A that gets as input a point $x \in \mathbb{R}^n$ and either

- ightharpoonup certifies that $x \in K$,
- or finds a hyperplane separating x from K.

We will usually assume that A is a polynomial-time algorithm.

In order to find a point in K we need

- ightharpoonup a guarantee that a ball of radius r is contained in K,
- ▶ an initial ball B(c,R) with radius R that contains K,
- ightharpoonup a separation oracle for K

Let $K \subseteq \mathbb{R}^n$ be a convex set. A separation oracle for K is an algorithm A that gets as input a point $x \in \mathbb{R}^n$ and either

- riangleright certifies that $x \in K$,
- or finds a hyperplane separating x from K.

We will usually assume that A is a polynomial-time algorithm.

In order to find a point in K we need

- ightharpoonup a guarantee that a ball of radius r is contained in K,
- ▶ an initial ball B(c,R) with radius R that contains K,
- ightharpoonup a separation oracle for K.

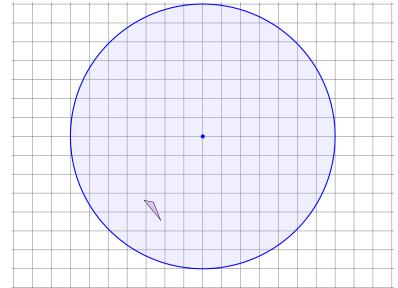
Let $K \subseteq \mathbb{R}^n$ be a convex set. A separation oracle for K is an algorithm A that gets as input a point $x \in \mathbb{R}^n$ and either

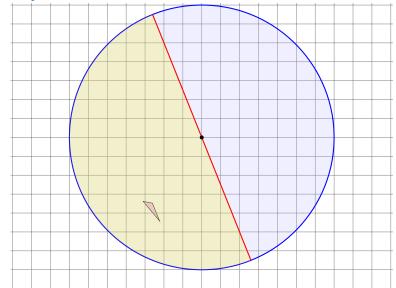
- ightharpoonup certifies that $x \in K$,
- or finds a hyperplane separating x from K.

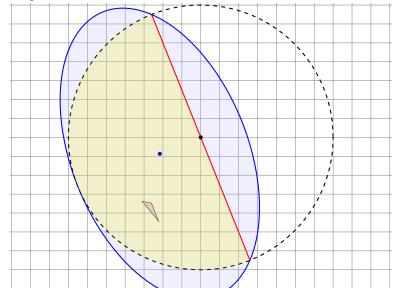
We will usually assume that A is a polynomial-time algorithm.

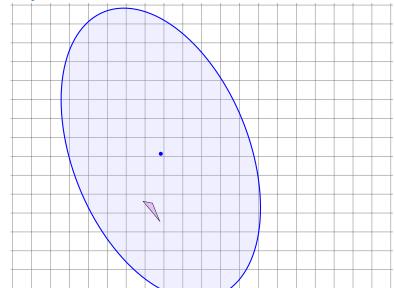
In order to find a point in K we need

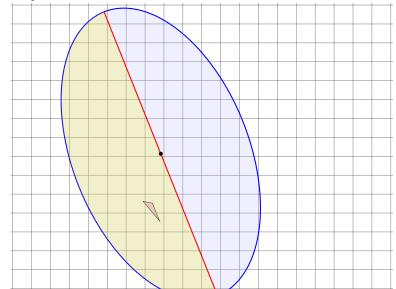
- ightharpoonup a guarantee that a ball of radius r is contained in K,
- ▶ an initial ball B(c,R) with radius R that contains K,
- ightharpoonup a separation oracle for K.

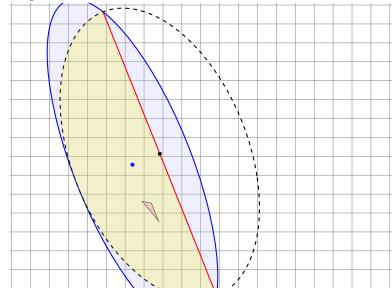


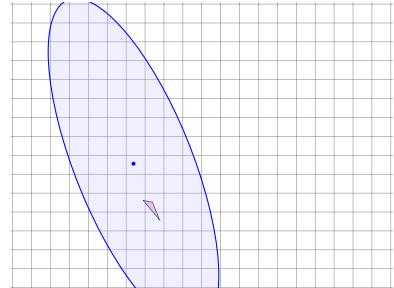


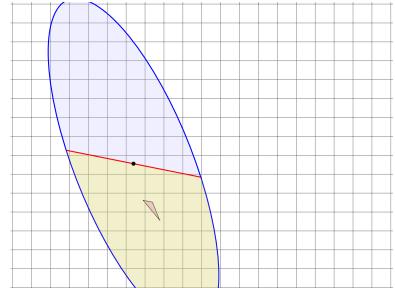


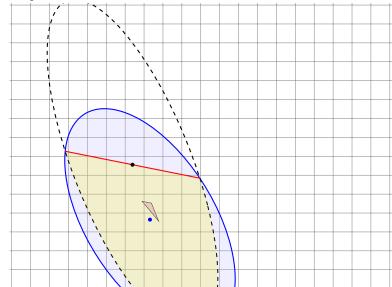


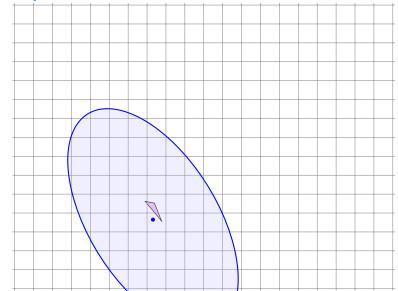


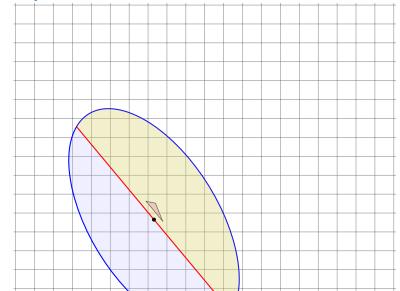


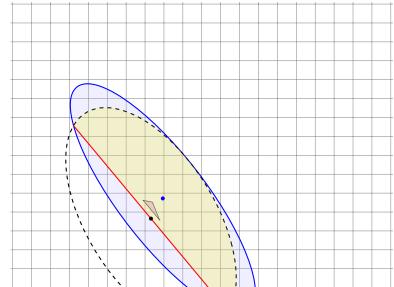


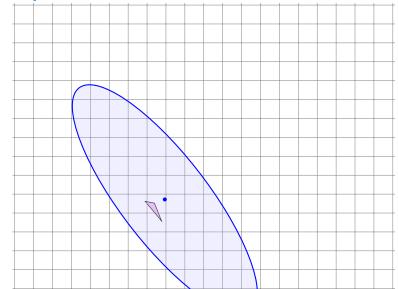


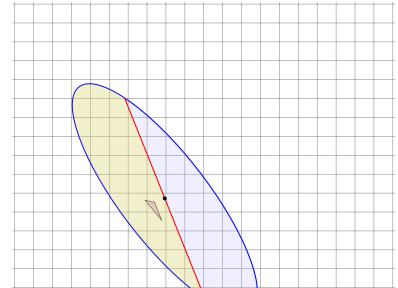


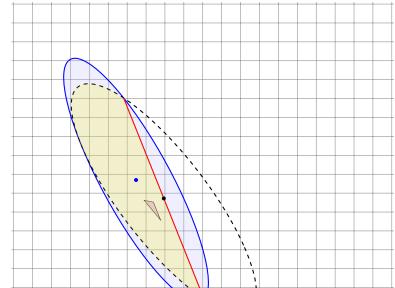


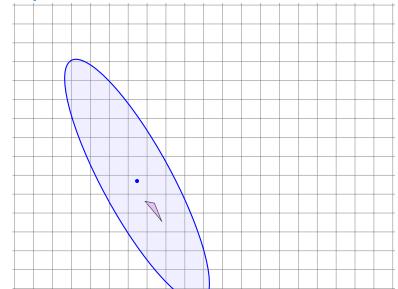


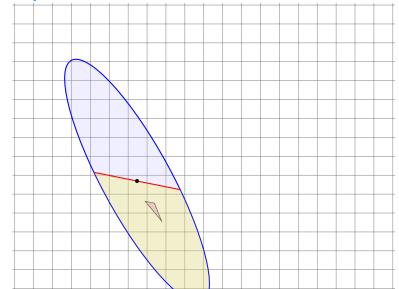


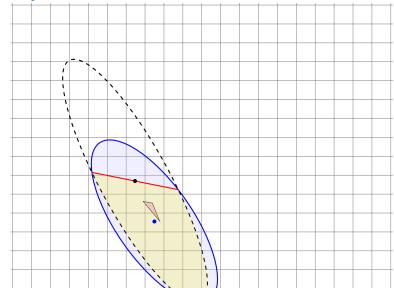


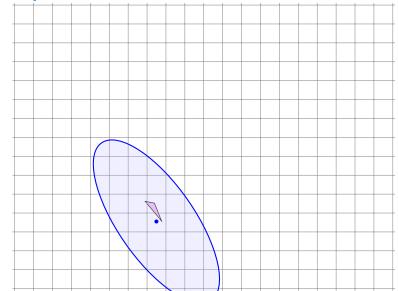


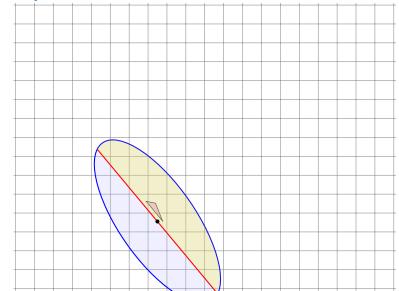


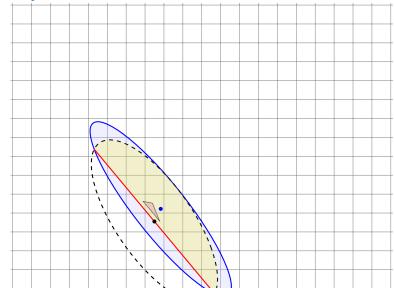


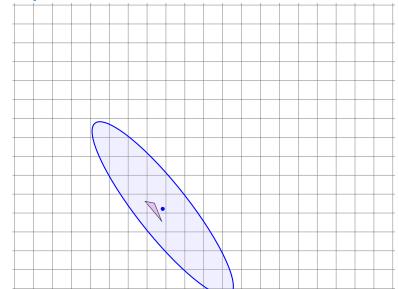


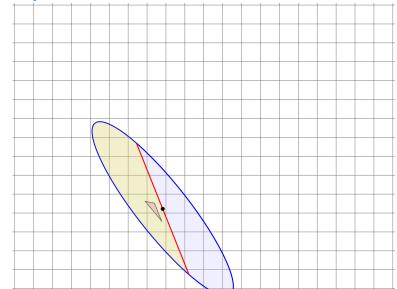


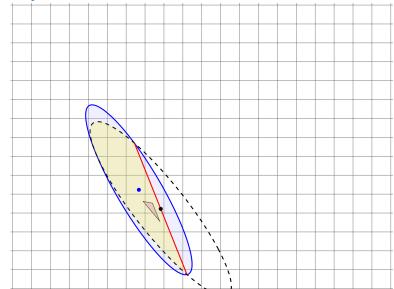


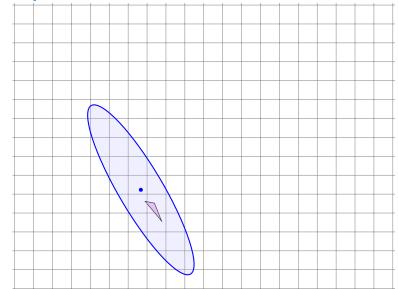


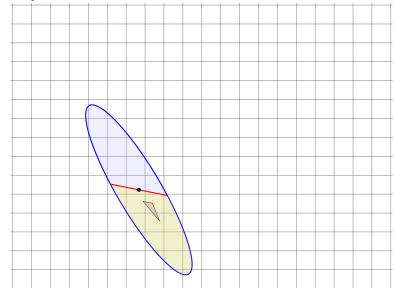


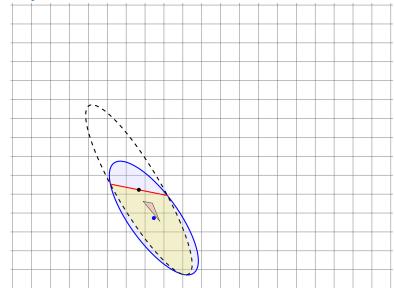


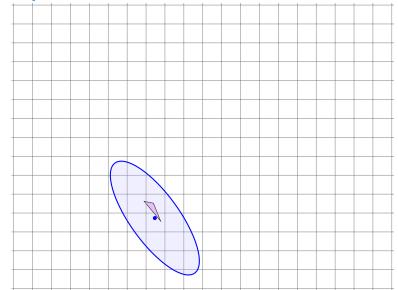


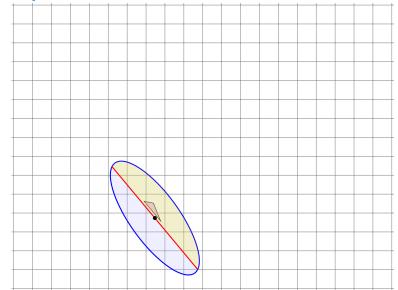


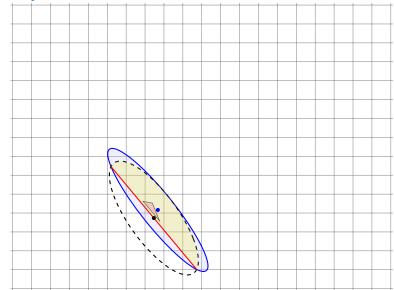


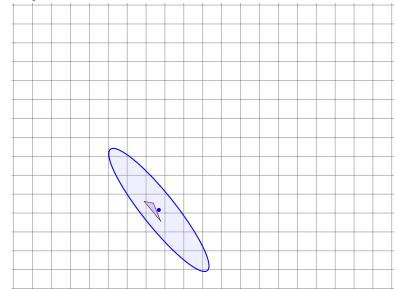


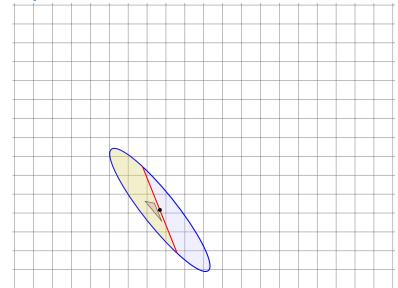


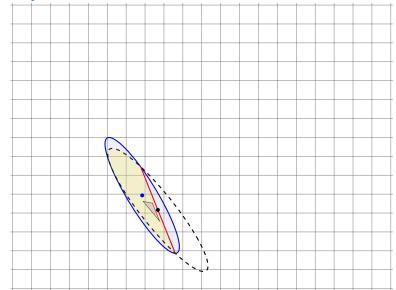


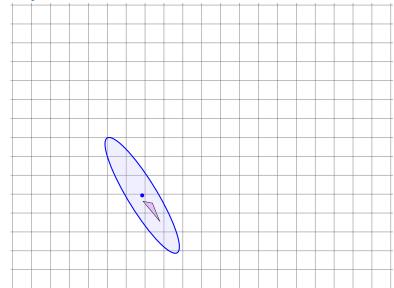


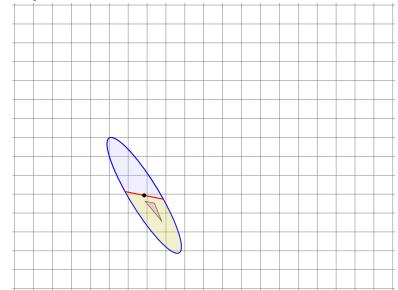


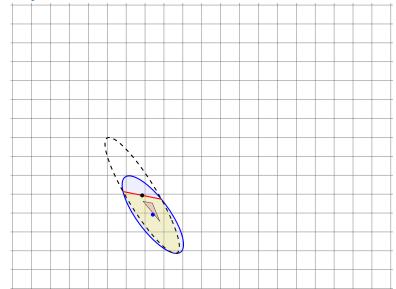


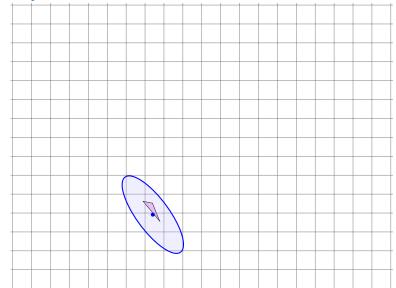


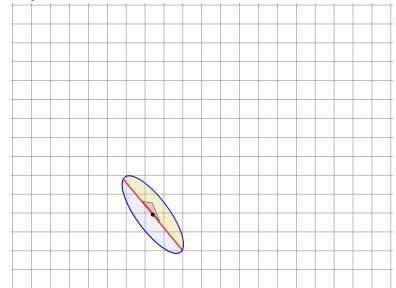


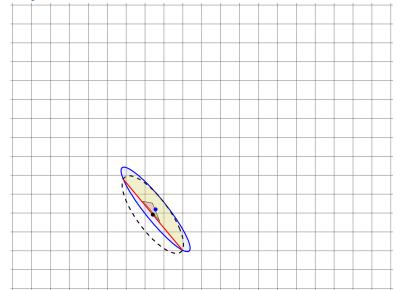


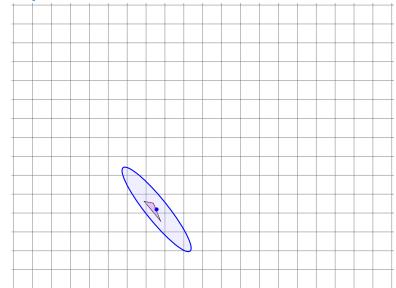


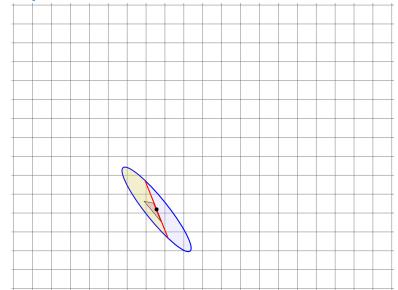


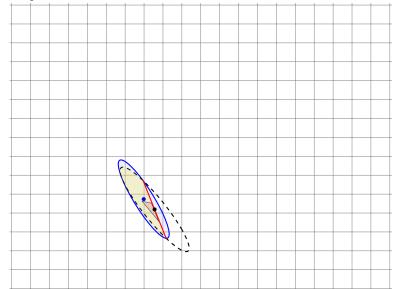


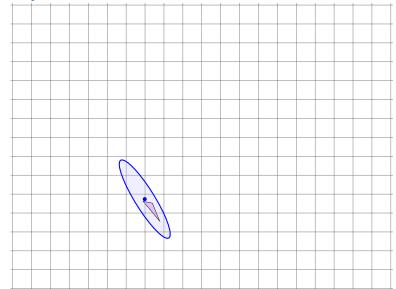


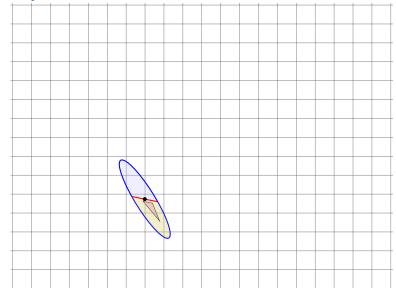


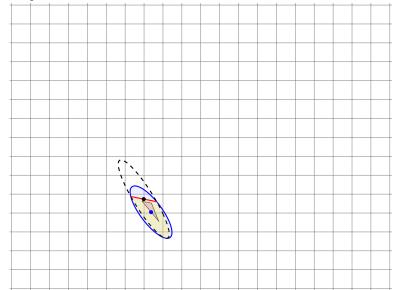


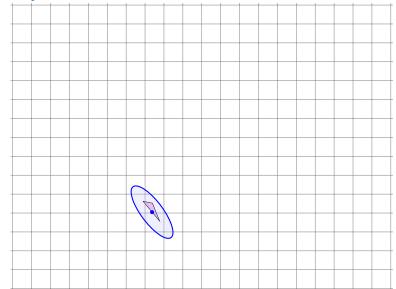


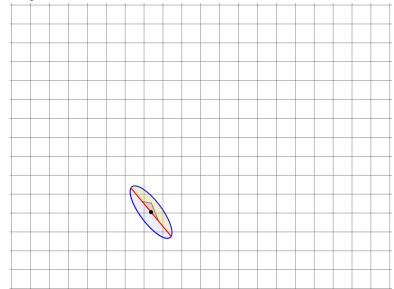


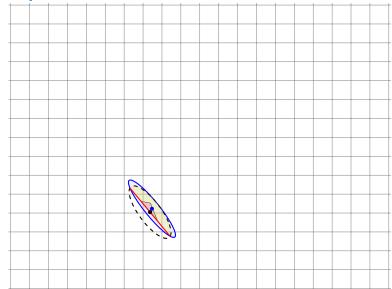


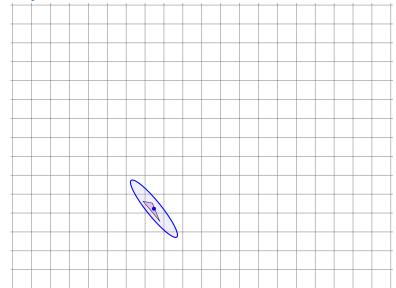












10 Karmarkars Algorithm

- ▶ inequalities $Ax \le b$; $m \times n$ matrix A with rows a_i^T
- $P = \{x \mid Ax \le b\}; P^{\circ} := \{x \mid Ax < b\}$
- lacktriangle interior point algorithm: $x\in P^\circ$ throughout the algorithm
- ▶ for $x \in P^\circ$ define

$$s_i(x) := b_i - a_i^T x$$

as the slack of the *i*-th constraint

logarithmic barrier function:

$$\phi(x) = -\sum_{i=1}^{m} \ln(s_i(x))$$

Penalty for point x; points close to the boundary have a very large penalty.

10 Karmarkars Algorithm

- ▶ inequalities $Ax \le b$; $m \times n$ matrix A with rows a_i^T
- ▶ $P = \{x \mid Ax \le b\}; P^{\circ} := \{x \mid Ax < b\}$
- ▶ interior point algorithm: $x \in P^{\circ}$ throughout the algorithm
- ▶ for $x \in P^\circ$ define

$$s_i(x) := b_i - a_i^T x$$

as the slack of the i-th constraint

logarithmic barrier function:

$$\phi(x) = -\sum_{i=1}^{m} \ln(s_i(x))$$

Penalty for point x; points close to the boundary have a very large penalty.

10 Karmarkars Algorithm

- ▶ inequalities $Ax \le b$; $m \times n$ matrix A with rows a_i^T
- ▶ $P = \{x \mid Ax \le b\}; P^{\circ} := \{x \mid Ax < b\}$
- ▶ interior point algorithm: $x \in P^{\circ}$ throughout the algorithm
- ▶ for $x \in P^\circ$ define

$$s_i(x) := b_i - a_i^T x$$

as the slack of the i-th constraint

logarithmic barrier function:

$$\phi(x) = -\sum_{i=1}^m \ln(s_i(x))$$

Penalty for point x; points close to the boundary have a very large penalty.

10 Karmarkars Algorithm

- ▶ inequalities $Ax \le b$; $m \times n$ matrix A with rows a_i^T
- ▶ $P = \{x \mid Ax \le b\}; P^{\circ} := \{x \mid Ax < b\}$
- ▶ interior point algorithm: $x \in P^{\circ}$ throughout the algorithm
- ▶ for $x \in P^{\circ}$ define

$$s_i(x) := b_i - a_i^T x$$

as the slack of the *i*-th constraint

logarithmic barrier function:

$$\phi(x) = -\sum_{i=1}^m \ln(s_i(x))$$

Penalty for point x; points close to the boundary have a very large penalty.

10 Karmarkars Algorithm

- ▶ inequalities $Ax \le b$; $m \times n$ matrix A with rows a_i^T
- ▶ $P = \{x \mid Ax \le b\}; P^{\circ} := \{x \mid Ax < b\}$
- ▶ interior point algorithm: $x \in P^{\circ}$ throughout the algorithm
- ▶ for $x \in P^{\circ}$ define

$$s_i(x) := b_i - a_i^T x$$

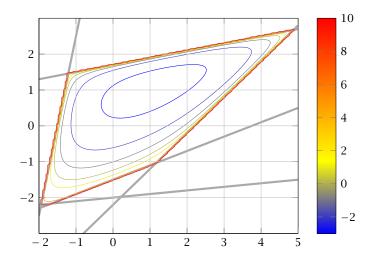
as the slack of the *i*-th constraint

logarithmic barrier function:

$$\phi(x) = -\sum_{i=1}^{m} \ln(s_i(x))$$

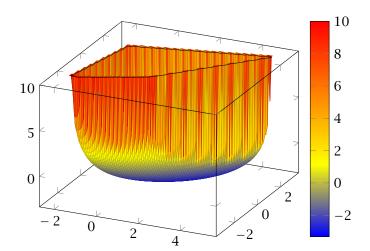
Penalty for point x; points close to the boundary have a very large penalty.

Penalty Function





Penalty Function





Gradient and Hessian

Taylor approximation:

$$\phi(x + \epsilon) \approx \phi(x) + \nabla \phi(x)^T \epsilon + \frac{1}{2} \epsilon^T \nabla^2 \phi(x) \epsilon$$

Gradient:

$$\nabla \phi(x) = \sum_{i=1}^{m} \frac{1}{s_i(x)} \cdot a_i = A^T d_X$$

where $d_x^T = (1/s_1(x), ..., 1/s_m(x))$. (d_x vector of inverse slacks)

Hessian:

$$H_X := \nabla^2 \phi(x) = \sum_{i=1}^m \frac{1}{s_i(x)^2} a_i a_i^T = A^T D_X^2 A_X^T$$

with $D_X = \operatorname{diag}(d_X)$.

Gradient and Hessian

Taylor approximation:

$$\phi(x + \epsilon) \approx \phi(x) + \nabla \phi(x)^T \epsilon + \frac{1}{2} \epsilon^T \nabla^2 \phi(x) \epsilon$$

Gradient:

$$\nabla \phi(x) = \sum_{i=1}^{m} \frac{1}{s_i(x)} \cdot a_i = A^T d_x$$

where $d_x^T = (1/s_1(x), ..., 1/s_m(x))$. (d_x vector of inverse slacks)

Hessian:

$$H_X := \nabla^2 \phi(x) = \sum_{i=1}^m \frac{1}{s_i(x)^2} a_i a_i^T = A^T D_x^2 A_x^T$$

with $D_X = \operatorname{diag}(d_X)$.

Gradient and Hessian

Taylor approximation:

$$\phi(x + \epsilon) \approx \phi(x) + \nabla \phi(x)^T \epsilon + \frac{1}{2} \epsilon^T \nabla^2 \phi(x) \epsilon$$

Gradient:

$$\nabla \phi(x) = \sum_{i=1}^{m} \frac{1}{s_i(x)} \cdot a_i = A^T d_x$$

where $d_x^T = (1/s_1(x), ..., 1/s_m(x))$. (d_x vector of inverse slacks)

Hessian:

$$H_X := \nabla^2 \phi(x) = \sum_{i=1}^m \frac{1}{s_i(x)^2} a_i a_i^T = A^T D_X^2 A$$

with $D_X = \operatorname{diag}(d_X)$.

Proof for Gradient

$$\begin{split} \frac{\partial \phi(x)}{\partial x_i} &= \frac{\partial}{\partial x_i} \left(-\sum_r \ln(s_r(x)) \right) \\ &= -\sum_r \frac{\partial}{\partial x_i} \left(\ln(s_r(x)) \right) = -\sum_r \frac{1}{s_r(x)} \frac{\partial}{\partial x_i} \left(s_r(x) \right) \\ &= -\sum_r \frac{1}{s_r(x)} \frac{\partial}{\partial x_i} \left(b_r - a_r^T x \right) = \sum_r \frac{1}{s_r(x)} \frac{\partial}{\partial x_i} \left(a_r^T x \right) \\ &= \sum_r \frac{1}{s_r(x)} A_{ri} \end{split}$$

The *i*-th entry of the gradient vector is $\sum_{r} 1/s_r(x) \cdot A_{ri}$. This gives that the gradient is

$$\nabla \phi(x) = \sum_{r} 1/s_r(x) a_r = A^T d_X$$

Proof for Hessian

$$\frac{\partial}{\partial x_j} \left(\sum_r \frac{1}{s_r(x)} A_{ri} \right) = \sum_r A_{ri} \left(-\frac{1}{s_r(x)^2} \right) \cdot \frac{\partial}{\partial x_j} \left(s_r(x) \right)$$
$$= \sum_r A_{ri} \frac{1}{s_r(x)^2} A_{rj}$$

Note that $\sum_r A_{ri} A_{rj} = (A^T A)_{ij}$. Adding the additional factors $1/s_r(x)^2$ can be done with a diagonal matrix.

Hence the Hessian is

$$H_{\mathcal{X}} = A^T D^2 A$$

 H_X is positive semi-definite for $X \in P^{\circ}$

$$u^{T}H_{x}u = u^{T}A^{T}D_{x}^{2}Au = ||D_{x}Au||_{2}^{2} \ge 0$$

This gives that $\phi(x)$ is convex

If rank(A) = n, H_X is positive definite for $X \in P^{\circ}$

$$u^T H_X u = ||D_X A u||_2^2 > 0 \text{ for } u \neq 0$$

This gives that $\phi(x)$ is strictly convex.

 H_X is positive semi-definite for $X \in P^{\circ}$

$$u^{T}H_{x}u = u^{T}A^{T}D_{x}^{2}Au = ||D_{x}Au||_{2}^{2} \ge 0$$

This gives that $\phi(x)$ is convex.

If rank(A) = n, H_X is positive definite for $X \in P^{\circ}$

$$u^{T}H_{X}u = \|D_{X}Au\|_{2}^{2} > 0 \text{ for } u \neq 0$$

This gives that $\phi(x)$ is strictly convex.

 H_X is positive semi-definite for $X \in P^{\circ}$

$$u^{T}H_{x}u = u^{T}A^{T}D_{x}^{2}Au = ||D_{x}Au||_{2}^{2} \ge 0$$

This gives that $\phi(x)$ is convex.

If
$$rank(A) = n$$
, H_X is positive definite for $X \in P^{\circ}$

$$u^{T}H_{x}u = ||D_{x}Au||_{2}^{2} > 0 \text{ for } u \neq 0$$

This gives that $\phi(x)$ is strictly convex.

 $H_{\mathcal{X}}$ is positive semi-definite for $x \in P^{\circ}$

$$u^{T}H_{x}u = u^{T}A^{T}D_{x}^{2}Au = ||D_{x}Au||_{2}^{2} \ge 0$$

This gives that $\phi(x)$ is convex.

If rank(A) = n, H_X is positive definite for $X \in P^{\circ}$

$$u^{T}H_{x}u = ||D_{x}Au||_{2}^{2} > 0 \text{ for } u \neq 0$$

This gives that $\phi(x)$ is strictly convex.

 H_X is positive semi-definite for $X \in P^\circ$

$$u^{T}H_{x}u = u^{T}A^{T}D_{x}^{2}Au = ||D_{x}Au||_{2}^{2} \ge 0$$

This gives that $\phi(x)$ is convex.

If rank(A) = n, H_X is positive definite for $X \in P^{\circ}$

$$u^T H_X u = ||D_X A u||_2^2 > 0 \text{ for } u \neq 0$$

This gives that $\phi(x)$ is strictly convex.

$$E_x = \{ y \mid (y - x)^T H_x (y - x) \le 1 \} = \{ y \mid ||y - x||_{H_x} \le 1 \}$$

Points in E_x are feasible!!!

change of distance to *i*-th constraint going from x to y distance of x to *i*-th constraint:

$$E_X = \{ y \mid (y - x)^T H_X (y - x) \le 1 \} = \{ y \mid ||y - x||_{H_X} \le 1 \}$$

Points in E_x are feasible!!!

$$(y - x)^{T} H_{X}(y - x) = (y - x)^{T} A^{T} D_{X}^{2} A(y - x)$$

$$= \sum_{i=1}^{m} \frac{(a_{i}^{T} (y - x))^{2}}{s_{i}(x)^{2}}$$

$$= \sum_{i=1}^{m} \frac{(\text{change of distance to } i\text{-th constraint going from } x \text{ to } y)^{2}}{(\text{distance of } x \text{ to } i\text{-th constraint})^{2}}$$

$$\leq 1$$

$$E_X = \{ y \mid (y - x)^T H_X(y - x) \le 1 \} = \{ y \mid ||y - x||_{H_X} \le 1 \}$$

Points in E_x are feasible!!!

$$(y-x)^{T}H_{x}(y-x) = (y-x)^{T}A^{T}D_{x}^{2}A(y-x)$$

$$= \sum_{i=1}^{m} \frac{(a_{i}^{T}(y-x))^{2}}{s_{i}(x)^{2}}$$

$$= \sum_{i=1}^{m} \frac{(\text{change of distance to } i\text{-th constraint going from } x \text{ to } y)^{2}}{(\text{distance of } x \text{ to } i\text{-th constraint})^{2}}$$

$$< 1$$

$$E_X = \{ y \mid (y - x)^T H_X (y - x) \le 1 \} = \{ y \mid ||y - x||_{H_X} \le 1 \}$$

Points in E_x are feasible!!!

$$(y-x)^{T}H_{x}(y-x) = (y-x)^{T}A^{T}D_{x}^{2}A(y-x)$$

$$= \sum_{i=1}^{m} \frac{(a_{i}^{T}(y-x))^{2}}{s_{i}(x)^{2}}$$

$$= \sum_{i=1}^{m} \frac{(\text{change of distance to } i\text{-th constraint going from } x \text{ to } y)^{2}}{(\text{distance of } x \text{ to } i\text{-th constraint})^{2}}$$

$$E_X = \{ y \mid (y - x)^T H_X (y - x) \le 1 \} = \{ y \mid ||y - x||_{H_X} \le 1 \}$$

Points in E_x are feasible!!!

$$(y - x)^{T} H_{X}(y - x) = (y - x)^{T} A^{T} D_{X}^{2} A(y - x)$$

$$= \sum_{i=1}^{m} \frac{(a_{i}^{T} (y - x))^{2}}{s_{i}(x)^{2}}$$

$$= \sum_{i=1}^{m} \frac{(\text{change of distance to } i\text{-th constraint going from } x \text{ to } y)^{2}}{(\text{distance of } x \text{ to } i\text{-th constraint})^{2}}$$

$$E_X = \{ y \mid (y - x)^T H_X (y - x) \le 1 \} = \{ y \mid ||y - x||_{H_X} \le 1 \}$$

Points in E_x are feasible!!!

$$(y - x)^{T} H_{X}(y - x) = (y - x)^{T} A^{T} D_{X}^{2} A(y - x)$$

$$= \sum_{i=1}^{m} \frac{(a_{i}^{T} (y - x))^{2}}{s_{i}(x)^{2}}$$

$$= \sum_{i=1}^{m} \frac{(\text{change of distance to } i\text{-th constraint going from } x \text{ to } y)^{2}}{(\text{distance of } x \text{ to } i\text{-th constraint})^{2}}$$

$$\leq 1$$

$$E_X = \{ y \mid (y - x)^T H_X (y - x) \le 1 \} = \{ y \mid ||y - x||_{H_X} \le 1 \}$$

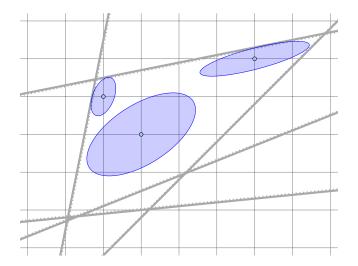
Points in E_x are feasible!!!

$$(y - x)^{T} H_{X}(y - x) = (y - x)^{T} A^{T} D_{X}^{2} A(y - x)$$

$$= \sum_{i=1}^{m} \frac{(a_{i}^{T} (y - x))^{2}}{s_{i}(x)^{2}}$$

$$= \sum_{i=1}^{m} \frac{(\text{change of distance to } i\text{-th constraint going from } x \text{ to } y)^{2}}{(\text{distance of } x \text{ to } i\text{-th constraint})^{2}}$$

$$< 1$$



Analytic Center

$$x_{\mathrm{ac}} := \operatorname{arg\,min}_{x \in P^{\circ}} \phi(x)$$

 \triangleright $x_{\rm ac}$ is solution to

$$\nabla \phi(x) = \sum_{i=1}^{m} \frac{1}{s_i(x)} a_i = 0$$

- depends on the description of the polytope
- \blacktriangleright $x_{\rm ac}$ exists and is unique iff P° is nonempty and bounded

In the following we assume that the LP and its dual are strictly feasible and that rank(A) = n.

```
Set of points \{x^*(t) \mid t > 0\} with x^*(t) = \operatorname{argmin}_{\sim} \{tc^Tx + \phi(x)\}
```

- t = 0: analytic center
- ► $t = \infty$: optimum solution
- $x^*(t)$ exists and is unique for all $t \ge 0$

In the following we assume that the LP and its dual are strictly feasible and that rank(A) = n.

Central Path:

Set of points $\{x^*(t) \mid t > 0\}$ with

$$x^*(t) = \operatorname{argmin}_{x} \{ tc^T x + \phi(x) \}$$

- t = 0: analytic center
- $t = \infty$: optimum solution

 $x^*(t)$ exists and is unique for all $t \ge 0$.

In the following we assume that the LP and its dual are strictly feasible and that rank(A) = n.

Central Path:

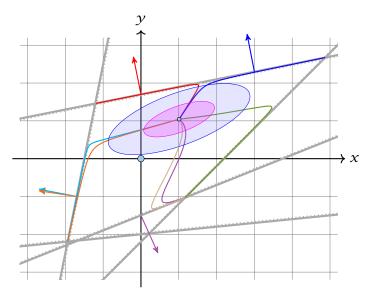
Set of points $\{x^*(t) \mid t > 0\}$ with

$$x^*(t) = \operatorname{argmin}_{x} \{ tc^T x + \phi(x) \}$$

- t = 0: analytic center
- ▶ $t = \infty$: optimum solution

 $x^*(t)$ exists and is unique for all $t \ge 0$.

Different Central Paths





Intuitive Idea:

Find point on central path for large value of t. Should be close to optimum solution.

Questions:

- Is this really true? How large a t do we need?
- ► How do we find corresponding point $x^*(t)$ on central path?

The Dual

primal-dual pair:

$$\begin{array}{ll}
\text{min } c^T x \\
\text{s.t. } Ax \leq b
\end{array}$$

$$\max -b^{T}z$$
s.t. $A^{T}z + c = 0$
 $z \ge 0$

Assumptions

- primal and dual problems are strictly feasible;
- ightharpoonup rank(A) = n.

Force Field Interpretation

Point $x^*(t)$ on central path is solution to $tc + \nabla \phi(x) = 0$

- We can view each constraint as generating a repelling force. The combination of these forces is represented by $\nabla \phi(x)$.
- In addition there is a force tc pulling us towards the optimum solution.

Point $x^*(t)$ on central path is solution to $tc + \nabla \phi(x) = 0$.

This means

$$tc + \sum_{i=1}^{m} \frac{1}{s_i(x^*(t))} a_i = 0$$

0ľ

$$c + \sum_{i=1}^{m} z_i^*(t) a_i = 0$$
 with $z_i^*(t) = \frac{1}{t s_i(x^*(t))}$

duality gap between a second and a second second

Point $x^*(t)$ on central path is solution to $tc + \nabla \phi(x) = 0$.

This means

$$tc + \sum_{i=1}^{m} \frac{1}{s_i(x^*(t))} a_i = 0$$

or

$$c + \sum_{i=1}^{m} z_i^*(t) a_i = 0$$
 with $z_i^*(t) = \frac{1}{t s_i(x^*(t))}$

a third strictly dual feasible: ((a)

duality gap between a second and a second is

Point $x^*(t)$ on central path is solution to $tc + \nabla \phi(x) = 0$.

This means

$$tc + \sum_{i=1}^{m} \frac{1}{s_i(x^*(t))} a_i = 0$$

or

$$c + \sum_{i=1}^{m} z_i^*(t) a_i = 0$$
 with $z_i^*(t) = \frac{1}{t s_i(x^*(t))}$

Point $x^*(t)$ on central path is solution to $tc + \nabla \phi(x) = 0$.

This means

$$tc + \sum_{i=1}^{m} \frac{1}{s_i(x^*(t))} a_i = 0$$

or

$$c + \sum_{i=1}^{m} z_i^*(t) a_i = 0$$
 with $z_i^*(t) = \frac{1}{t s_i(x^*(t))}$

- \triangleright $z^*(t)$ is strictly dual feasible: $(A^Tz^* + c = 0; z^* > 0)$
- duality gap between $x := x^*(t)$ and $z := z^*(t)$ is

$$c^T x + b^T z = (b - Ax)^T z = \frac{m}{t}$$

• if gap is less than $1/2^{\Omega(L)}$ we can snap to optimum point

Point $x^*(t)$ on central path is solution to $tc + \nabla \phi(x) = 0$.

This means

$$tc + \sum_{i=1}^{m} \frac{1}{s_i(x^*(t))} a_i = 0$$

or

$$c + \sum_{i=1}^{m} z_i^*(t) a_i = 0$$
 with $z_i^*(t) = \frac{1}{t s_i(x^*(t))}$

- \triangleright $z^*(t)$ is strictly dual feasible: $(A^Tz^* + c = 0; z^* > 0)$
- duality gap between $x := x^*(t)$ and $z := z^*(t)$ is

$$c^T x + b^T z = (b - Ax)^T z = \frac{m}{t}$$

• if gap is less than $1/2^{\Omega(L)}$ we can snap to optimum point

Point $x^*(t)$ on central path is solution to $tc + \nabla \phi(x) = 0$.

This means

$$tc + \sum_{i=1}^{m} \frac{1}{s_i(x^*(t))} a_i = 0$$

or

$$c + \sum_{i=1}^{m} z_i^*(t) a_i = 0$$
 with $z_i^*(t) = \frac{1}{t s_i(x^*(t))}$

- \triangleright $z^*(t)$ is strictly dual feasible: $(A^Tz^* + c = 0; z^* > 0)$
- duality gap between $x := x^*(t)$ and $z := z^*(t)$ is

$$c^T x + b^T z = (b - Ax)^T z = \frac{m}{t}$$

• if gap is less than $1/2^{\Omega(L)}$ we can snap to optimum point

How to find $x^*(t)$

First idea:

- start somewhere in the polytope
- use iterative method (Newtons method) to minimize $f_t(x) := tc^T x + \phi(x)$

Quadratic approximation of f_t

$$f_t(x + \epsilon) \approx f_t(x) + \nabla f_t(x)^T \epsilon + \frac{1}{2} \epsilon^T H_{f_t}(x) \, \epsilon$$

Suppose this were exact:

$$f_t(x + \epsilon) = f_t(x) + \nabla f_t(x)^T \epsilon + \frac{1}{2} \epsilon^T H_{f_t}(x) \epsilon$$

Then gradient is given by:

$$\nabla f_t(x + \epsilon) = \nabla f_t(x) + H_{f_t}(x) \cdot \epsilon$$

Quadratic approximation of f_t

$$f_t(x + \epsilon) \approx f_t(x) + \nabla f_t(x)^T \epsilon + \frac{1}{2} \epsilon^T H_{f_t}(x) \epsilon$$

Suppose this were exact:

$$f_t(x + \epsilon) = f_t(x) + \nabla f_t(x)^T \epsilon + \frac{1}{2} \epsilon^T H_{f_t}(x) \epsilon$$

Then gradient is given by:

$$\nabla f_t(x + \epsilon) = \nabla f_t(x) + H_{f_t}(x) \cdot \epsilon$$

Quadratic approximation of f_t

$$f_t(x + \epsilon) \approx f_t(x) + \nabla f_t(x)^T \epsilon + \frac{1}{2} \epsilon^T H_{f_t}(x) \epsilon$$

Suppose this were exact:

$$f_t(x + \epsilon) = f_t(x) + \nabla f_t(x)^T \epsilon + \frac{1}{2} \epsilon^T H_{f_t}(x) \epsilon$$

Then gradient is given by:

$$\nabla f_t(x + \epsilon) = \nabla f_t(x) + H_{f_t}(x) \cdot \epsilon$$

We want to move to a point where this gradient is 0:

Newton Step at $x \in P^{\circ}$

$$\begin{split} \Delta x_{\mathsf{nt}} &= -H_{ft}^{-1}(x) \nabla f_t(x) \\ &= -H_{ft}^{-1}(x) (tc + \nabla \phi(x)) \\ &= -(A^T D_x^2 A)^{-1} (tc + A^T d_x) \end{split}$$

Newton Iteration:

$$x := x + \Delta x_{nt}$$

Measuring Progress of Newton Step

Newton decrement:

$$\lambda_t(x) = \|D_x A \Delta x_{\mathsf{nt}}\|$$
$$= \|\Delta x_{\mathsf{nt}}\|_{H_x}$$

Square of Newton decrement is linear estimate of reduction if we do a Newton step:

$$-\lambda_t(x)^2 = \nabla f_t(x)^T \Delta x_{\mathsf{nt}}$$

- $\lambda_t(x) = 0 \text{ iff } x = x^*(t)$
- $ightharpoonup \lambda_t(x)$ is measure of proximity of x to $x^*(t)$

Measuring Progress of Newton Step

Newton decrement:

$$\lambda_t(x) = \|D_x A \Delta x_{\mathsf{nt}}\|$$
$$= \|\Delta x_{\mathsf{nt}}\|_{H_x}$$

Square of Newton decrement is linear estimate of reduction if we do a Newton step:

$$-\lambda_t(x)^2 = \nabla f_t(x)^T \Delta x_{\mathsf{nt}}$$

- $\lambda_t(x) = 0 \text{ iff } x = x^*(t)$
- $ightharpoonup \lambda_t(x)$ is measure of proximity of x to $x^*(t)$

Measuring Progress of Newton Step

Newton decrement:

$$\lambda_t(x) = \|D_x A \Delta x_{\mathsf{nt}}\|$$
$$= \|\Delta x_{\mathsf{nt}}\|_{H_x}$$

Square of Newton decrement is linear estimate of reduction if we do a Newton step:

$$-\lambda_t(x)^2 = \nabla f_t(x)^T \Delta x_{\mathsf{nt}}$$

- $\lambda_t(x)$ is measure of proximity of x to $x^*(t)$

Theorem 55

If $\lambda_t(x) < 1$ then

- $x_+ := x + \Delta x_{nt} \in P^\circ$ (new point feasible)
- $\lambda_t(x_+) \leq \lambda_t(x)^2$

This means we have quadratic convergence. Very fast.

feasibility:

 $\lambda_t(x) = \|\Delta x_{\mathsf{nt}}\|_{H_x} < 1$; hence x_+ lies in the Dikin ellipsoid around x.

bound on $\lambda_t(x^+)$:

we use
$$D := D_X = \operatorname{diag}(d_X)$$
 and $D_+ := D_{X^+} = \operatorname{diag}(d_{X^+})$

To see the last equality we use Pythagoras

$$||a||^2 + ||a + b||^2 = ||b||^2$$

if
$$a^T(a+b)=0$$

bound on $\lambda_t(x^+)$:

we use
$$D := D_X = \operatorname{diag}(d_X)$$
 and $D_+ := D_{X^+} = \operatorname{diag}(d_{X^+})$

$$\lambda_{t}(x^{+})^{2} = \|D_{+}A\Delta x_{nt}^{+}\|^{2}$$

$$\leq \|D_{+}A\Delta x_{nt}^{+}\|^{2} + \|D_{+}A\Delta x_{nt}^{+} + (I - D_{+}^{-1}D)DA\Delta x_{nt}\|^{2}$$

$$= \|(I - D_{+}^{-1}D)DA\Delta x_{nt}\|^{2}$$

To see the last equality we use Pythagora:

$$||a||^2 + ||a + b||^2 = ||b||^2$$

if
$$a^T(a+b) = 0$$
.

bound on $\lambda_t(x^+)$:

we use
$$D := D_X = \text{diag}(d_X)$$
 and $D_+ := D_{X^+} = \text{diag}(d_{X^+})$

$$\lambda_{t}(x^{+})^{2} = \|D_{+}A\Delta x_{\mathsf{nt}}^{+}\|^{2}$$

$$\leq \|D_{+}A\Delta x_{\mathsf{nt}}^{+}\|^{2} + \|D_{+}A\Delta x_{\mathsf{nt}}^{+} + (I - D_{+}^{-1}D)DA\Delta x_{\mathsf{nt}}\|^{2}$$

$$= \|(I - D_{+}^{-1}D)DA\Delta x_{\mathsf{nt}}\|^{2}$$

To see the last equality we use Pythagoras

$$||a||^2 + ||a + b||^2 = ||b||^2$$

if
$$a^T(a+b)=0$$

bound on $\lambda_t(x^+)$:

we use
$$D := D_X = \operatorname{diag}(d_X)$$
 and $D_+ := D_{X^+} = \operatorname{diag}(d_{X^+})$

$$\lambda_{t}(x^{+})^{2} = \|D_{+}A\Delta x_{\mathsf{nt}}^{+}\|^{2}$$

$$\leq \|D_{+}A\Delta x_{\mathsf{nt}}^{+}\|^{2} + \|D_{+}A\Delta x_{\mathsf{nt}}^{+} + (I - D_{+}^{-1}D)DA\Delta x_{\mathsf{nt}}\|^{2}$$

$$= \|(I - D_{+}^{-1}D)DA\Delta x_{\mathsf{nt}}\|^{2}$$

To see the last equality we use Pythagora:

$$||a||^2 + ||a + b||^2 = ||b||^2$$

if
$$a^T(a+b) = 0$$
.

bound on $\lambda_t(x^+)$:

we use
$$D := D_X = \text{diag}(d_X)$$
 and $D_+ := D_{X^+} = \text{diag}(d_{X^+})$

$$\lambda_{t}(x^{+})^{2} = \|D_{+}A\Delta x_{\mathsf{nt}}^{+}\|^{2}$$

$$\leq \|D_{+}A\Delta x_{\mathsf{nt}}^{+}\|^{2} + \|D_{+}A\Delta x_{\mathsf{nt}}^{+} + (I - D_{+}^{-1}D)DA\Delta x_{\mathsf{nt}}\|^{2}$$

$$= \|(I - D_{+}^{-1}D)DA\Delta x_{\mathsf{nt}}\|^{2}$$

To see the last equality we use Pythagoras

$$||a||^2 + ||a + b||^2 = ||b||^2$$

if
$$a^T(a+b) = 0$$
.

bound on $\lambda_t(x^+)$:

we use
$$D := D_X = \text{diag}(d_X)$$
 and $D_+ := D_{X^+} = \text{diag}(d_{X^+})$

$$\lambda_{t}(x^{+})^{2} = \|D_{+}A\Delta x_{\mathsf{nt}}^{+}\|^{2}$$

$$\leq \|D_{+}A\Delta x_{\mathsf{nt}}^{+}\|^{2} + \|D_{+}A\Delta x_{\mathsf{nt}}^{+} + (I - D_{+}^{-1}D)DA\Delta x_{\mathsf{nt}}\|^{2}$$

$$= \|(I - D_{+}^{-1}D)DA\Delta x_{\mathsf{nt}}\|^{2}$$

To see the last equality we use Pythagoras

$$||a||^2 + ||a + b||^2 = ||b||^2$$

if
$$a^T(a+b)=0$$
.

$$DA\Delta x_{\mathsf{nt}} = DA(x^{+} - x)$$

$$= D(b - Ax - (b - Ax^{+}))$$

$$= D(D^{-1}\vec{1} - D_{+}^{-1}\vec{1})$$

$$= (I - D_{-}^{-1}D)\vec{1}$$

$$a^{T}(a+b)$$

$$= \Delta x_{\mathsf{nt}}^{+T} A^{T} D_{+} \left(D_{+} A \Delta x_{\mathsf{nt}}^{+} + (I - D_{+}^{-1} D) D A \Delta x_{\mathsf{nt}} \right)$$

$$= \Delta x_{\mathsf{nt}}^{+T} \left(A^{T} D_{+}^{2} A \Delta x_{\mathsf{nt}}^{+} - A^{T} D^{2} A \Delta x_{\mathsf{nt}} + A^{T} D_{+} D A \Delta x_{\mathsf{nt}} \right)$$

$$= \Delta x_{\mathsf{nt}}^{+T} \left(H_{+} \Delta x_{\mathsf{nt}}^{+} - H \Delta x_{\mathsf{nt}} + A^{T} D_{+} \vec{1} - A^{T} D \vec{1} \right)$$

$$= \Delta x_{\mathsf{nt}}^{+T} \left(- \nabla f_{t}(x^{+}) + \nabla f_{t}(x) + \nabla \phi(x^{+}) - \nabla \phi(x) \right)$$

$$= 0$$

$$DA\Delta x_{\mathsf{nt}} = DA(x^{+} - x)$$

$$= D(b - Ax - (b - Ax^{+}))$$

$$= D(D^{-1}\bar{1} - D_{+}^{-1}\bar{1})$$

$$= (I - D_{-}^{-1}D)\bar{1}$$

$$a^{T}(a+b)$$

$$= \Delta x_{\mathsf{nt}}^{+T} A^{T} D_{+} \left(D_{+} A \Delta x_{\mathsf{nt}}^{+} + (I - D_{+}^{-1} D) D A \Delta x_{\mathsf{nt}} \right)$$

$$= \Delta x_{\mathsf{nt}}^{+T} \left(A^{T} D_{+}^{2} A \Delta x_{\mathsf{nt}}^{+} - A^{T} D^{2} A \Delta x_{\mathsf{nt}} + A^{T} D_{+} D A \Delta x_{\mathsf{nt}} \right)$$

$$= \Delta x_{\mathsf{nt}}^{+T} \left(H_{+} \Delta x_{\mathsf{nt}}^{+} - H \Delta x_{\mathsf{nt}} + A^{T} D_{+} \vec{\mathbf{I}} - A^{T} D \vec{\mathbf{I}} \right)$$

$$= \Delta x_{\mathsf{nt}}^{+T} \left(- \nabla f_{t}(x^{+}) + \nabla f_{t}(x) + \nabla \phi(x^{+}) - \nabla \phi(x) \right)$$

$$= 0$$

$$DA\Delta x_{\mathsf{nt}} = DA(x^{+} - x)$$

$$= D(b - Ax - (b - Ax^{+}))$$

$$= D(D^{-1} \overline{1} - D_{+}^{-1} \overline{1})$$

$$= (I - D_{-}^{-1} D) \overline{1}$$

$$a^{T}(a+b)$$

$$= \Delta x_{\mathsf{nt}}^{+T} A^{T} D_{+} \left(D_{+} A \Delta x_{\mathsf{nt}}^{+} + (I - D_{+}^{-1} D) D A \Delta x_{\mathsf{nt}} \right)$$

$$= \Delta x_{\mathsf{nt}}^{+T} \left(A^{T} D_{+}^{2} A \Delta x_{\mathsf{nt}}^{+} - A^{T} D^{2} A \Delta x_{\mathsf{nt}} + A^{T} D_{+} D A \Delta x_{\mathsf{nt}} \right)$$

$$= \Delta x_{\mathsf{nt}}^{+T} \left(H_{+} \Delta x_{\mathsf{nt}}^{+} - H \Delta x_{\mathsf{nt}} + A^{T} D_{+} \vec{1} - A^{T} D \vec{1} \right)$$

$$= \Delta x_{\mathsf{nt}}^{+T} \left(- \nabla f_{t}(x^{+}) + \nabla f_{t}(x) + \nabla \phi(x^{+}) - \nabla \phi(x) \right)$$

$$= 0$$

$$DA\Delta x_{nt} = DA(x^{+} - x)$$

$$= D(b - Ax - (b - Ax^{+}))$$

$$= D(D^{-1}\vec{1} - D_{+}^{-1}\vec{1})$$

$$= (I - D_{-}^{-1}D)\vec{1}$$

$$a^{T}(a+b)$$

$$= \Delta x_{\mathsf{nt}}^{+T} A^{T} D_{+} \left(D_{+} A \Delta x_{\mathsf{nt}}^{+} + (I - D_{+}^{-1} D) D A \Delta x_{\mathsf{nt}} \right)$$

$$= \Delta x_{\mathsf{nt}}^{+T} \left(A^{T} D_{+}^{2} A \Delta x_{\mathsf{nt}}^{+} - A^{T} D^{2} A \Delta x_{\mathsf{nt}} + A^{T} D_{+} D A \Delta x_{\mathsf{nt}} \right)$$

$$= \Delta x_{\mathsf{nt}}^{+T} \left(H_{+} \Delta x_{\mathsf{nt}}^{+} - H \Delta x_{\mathsf{nt}} + A^{T} D_{+} \vec{\mathbf{I}} - A^{T} D \vec{\mathbf{I}} \right)$$

$$= \Delta x_{\mathsf{nt}}^{+T} \left(- \nabla f_{t}(x^{+}) + \nabla f_{t}(x) + \nabla \phi(x^{+}) - \nabla \phi(x) \right)$$

$$= 0$$

$$DA\Delta x_{\mathsf{nt}} = DA(x^{+} - x)$$

$$= D(b - Ax - (b - Ax^{+}))$$

$$= D(D^{-1}\vec{1} - D_{+}^{-1}\vec{1})$$

$$= (I - D_{+}^{-1}D)\vec{1}$$

$$a^{T}(a + b)$$

$$= \Delta x_{nt}^{+T} A^{T} D_{+} \left(D_{+} A \Delta x_{nt}^{+} + (I - D_{+}^{-1} D) D A \Delta x_{nt} \right)$$

$$= \Delta x_{nt}^{+T} \left(A^{T} D_{+}^{2} A \Delta x_{nt}^{+} - A^{T} D^{2} A \Delta x_{nt} + A^{T} D_{+} D A \Delta x_{nt} \right)$$

$$= \Delta x_{nt}^{+T} \left(H_{+} \Delta x_{nt}^{+} - H \Delta x_{nt} + A^{T} D_{+} \vec{1} - A^{T} D \vec{1} \right)$$

$$= \Delta x_{nt}^{+T} \left(- \nabla f_{t}(x^{+}) + \nabla f_{t}(x) + \nabla \phi(x^{+}) - \nabla \phi(x) \right)$$

$$= 0$$

$$DA\Delta x_{nt} = DA(x^{+} - x)$$

$$= D(b - Ax - (b - Ax^{+}))$$

$$= D(D^{-1}\vec{1} - D_{+}^{-1}\vec{1})$$

$$= (I - D_{+}^{-1}D)\vec{1}$$

$$a^{T}(a+b)$$

$$= \Delta x_{\text{nt}}^{+T} A^{T} D_{+} \left(D_{+} A \Delta x_{\text{nt}}^{+} + (I - D_{+}^{-1} D) D A \Delta x_{\text{nt}} \right)$$

$$= \Delta x_{\text{nt}}^{+T} \left(A^{T} D_{+}^{2} A \Delta x_{\text{nt}}^{+} - A^{T} D^{2} A \Delta x_{\text{nt}} + A^{T} D_{+} D A \Delta x_{\text{nt}} \right)$$

$$= \Delta x_{\text{nt}}^{+T} \left(H_{+} \Delta x_{\text{nt}}^{+} - H \Delta x_{\text{nt}} + A^{T} D_{+} \vec{\mathbf{I}} - A^{T} D \vec{\mathbf{I}} \right)$$

$$= \Delta x_{\text{nt}}^{+T} \left(- \nabla f_{t}(x^{+}) + \nabla f_{t}(x) + \nabla \phi(x^{+}) - \nabla \phi(x) \right)$$

$$= 0$$

$$DA\Delta x_{\mathsf{nt}} = DA(x^{+} - x)$$

$$= D(b - Ax - (b - Ax^{+}))$$

$$= D(D^{-1}\vec{1} - D_{+}^{-1}\vec{1})$$

$$= (I - D_{+}^{-1}D)\vec{1}$$

$$a^{T}(a+b)$$

$$= \Delta x_{\mathsf{nt}}^{+T} A^{T} D_{+} \left(D_{+} A \Delta x_{\mathsf{nt}}^{+} + (I - D_{+}^{-1} D) D A \Delta x_{\mathsf{nt}} \right)$$

$$= \Delta x_{\mathsf{nt}}^{+T} \left(A^{T} D_{+}^{2} A \Delta x_{\mathsf{nt}}^{+} - A^{T} D^{2} A \Delta x_{\mathsf{nt}} + A^{T} D_{+} D A \Delta x_{\mathsf{nt}} \right)$$

$$= \Delta x_{\mathsf{nt}}^{+T} \left(H_{+} \Delta x_{\mathsf{nt}}^{+} - H \Delta x_{\mathsf{nt}} + A^{T} D_{+} \overline{1} - A^{T} D \overline{1} \right)$$

$$= \Delta x_{\mathsf{nt}}^{+T} \left(- \nabla f_{t}(x^{+}) + \nabla f_{t}(x) + \nabla \phi(x^{+}) - \nabla \phi(x) \right)$$

$$= 0$$

$$DA\Delta x_{\mathsf{nt}} = DA(x^{+} - x)$$

$$= D(b - Ax - (b - Ax^{+}))$$

$$= D(D^{-1}\vec{1} - D_{+}^{-1}\vec{1})$$

$$= (I - D_{+}^{-1}D)\vec{1}$$

$$a^{T}(a+b)$$

$$= \Delta x_{\mathsf{nt}}^{+T} A^{T} D_{+} \left(D_{+} A \Delta x_{\mathsf{nt}}^{+} + (I - D_{+}^{-1} D) D A \Delta x_{\mathsf{nt}} \right)$$

$$= \Delta x_{\mathsf{nt}}^{+T} \left(A^{T} D_{+}^{2} A \Delta x_{\mathsf{nt}}^{+} - A^{T} D^{2} A \Delta x_{\mathsf{nt}} + A^{T} D_{+} D A \Delta x_{\mathsf{nt}} \right)$$

$$= \Delta x_{\mathsf{nt}}^{+T} \left(H_{+} \Delta x_{\mathsf{nt}}^{+} - H \Delta x_{\mathsf{nt}} + A^{T} D_{+} \mathbf{I} - A^{T} D \mathbf{I} \right)$$

$$= \Delta x_{\mathsf{nt}}^{+T} \left(- \nabla f_{t}(x^{+}) + \nabla f_{t}(x) + \nabla \phi(x^{+}) - \nabla \phi(x) \right)$$

$$= 0$$

$$DA\Delta x_{nt} = DA(x^{+} - x)$$

$$= D(b - Ax - (b - Ax^{+}))$$

$$= D(D^{-1}\vec{1} - D_{+}^{-1}\vec{1})$$

$$= (I - D_{+}^{-1}D)\vec{1}$$

$$a^{T}(a+b)$$

$$= \Delta x_{\mathsf{nt}}^{+T} A^{T} D_{+} \left(D_{+} A \Delta x_{\mathsf{nt}}^{+} + (I - D_{+}^{-1} D) D A \Delta x_{\mathsf{nt}} \right)$$

$$= \Delta x_{\mathsf{nt}}^{+T} \left(A^{T} D_{+}^{2} A \Delta x_{\mathsf{nt}}^{+} - A^{T} D^{2} A \Delta x_{\mathsf{nt}} + A^{T} D_{+} D A \Delta x_{\mathsf{nt}} \right)$$

$$= \Delta x_{\mathsf{nt}}^{+T} \left(H_{+} \Delta x_{\mathsf{nt}}^{+} - H \Delta x_{\mathsf{nt}} + A^{T} D_{+} \vec{\mathbf{1}} - A^{T} D \vec{\mathbf{1}} \right)$$

$$= \Delta x_{\mathsf{nt}}^{+T} \left(- \nabla f_{I}(x^{+}) + \nabla f_{I}(x) + \nabla \phi(x^{+}) - \nabla \phi(x) \right)$$

$$= 0$$

$$DA\Delta x_{\mathsf{nt}} = DA(x^{+} - x)$$

$$= D(b - Ax - (b - Ax^{+}))$$

$$= D(D^{-1}\vec{1} - D_{+}^{-1}\vec{1})$$

$$= (I - D_{+}^{-1}D)\vec{1}$$

$$a^{T}(a+b)$$

$$= \Delta x_{\mathsf{nt}}^{+T} A^{T} D_{+} \left(D_{+} A \Delta x_{\mathsf{nt}}^{+} + (I - D_{+}^{-1} D) D A \Delta x_{\mathsf{nt}} \right)$$

$$= \Delta x_{\mathsf{nt}}^{+T} \left(A^{T} D_{+}^{2} A \Delta x_{\mathsf{nt}}^{+} - A^{T} D^{2} A \Delta x_{\mathsf{nt}} + A^{T} D_{+} D A \Delta x_{\mathsf{nt}} \right)$$

$$= \Delta x_{\mathsf{nt}}^{+T} \left(H_{+} \Delta x_{\mathsf{nt}}^{+} - H \Delta x_{\mathsf{nt}} + A^{T} D_{+} \vec{1} - A^{T} D \vec{1} \right)$$

$$= \Delta x_{\mathsf{nt}}^{+T} \left(- \nabla f_{t}(x^{+}) + \nabla f_{t}(x) + \nabla \phi(x^{+}) - \nabla \phi(x) \right)$$

$$DA\Delta x_{nt} = DA(x^{+} - x)$$

$$= D(b - Ax - (b - Ax^{+}))$$

$$= D(D^{-1}\vec{1} - D_{+}^{-1}\vec{1})$$

$$= (I - D_{+}^{-1}D)\vec{1}$$

$$\begin{split} a^T(a+b) \\ &= \Delta x_{\mathsf{nt}}^{+T} A^T D_+ \left(D_+ A \Delta x_{\mathsf{nt}}^+ + (I - D_+^{-1} D) D A \Delta x_{\mathsf{nt}} \right) \\ &= \Delta x_{\mathsf{nt}}^{+T} \left(A^T D_+^2 A \Delta x_{\mathsf{nt}}^+ - A^T D^2 A \Delta x_{\mathsf{nt}} + A^T D_+ D A \Delta x_{\mathsf{nt}} \right) \\ &= \Delta x_{\mathsf{nt}}^{+T} \left(H_+ \Delta x_{\mathsf{nt}}^+ - H \Delta x_{\mathsf{nt}} + A^T D_+ \vec{\mathbf{I}} - A^T D \vec{\mathbf{I}} \right) \\ &= \Delta x_{\mathsf{nt}}^{+T} \left(- \nabla f_t(x^+) + \nabla f_t(x) + \nabla \phi(x^+) - \nabla \phi(x) \right) \\ &= 0 \end{split}$$

bound on $\lambda_t(x^+)$: we use $D := D_x = \operatorname{diag}(d_x)$ and $D_+ := D_{x^+} = \operatorname{diag}(d_{x^+})$ $\lambda_t(x^+)^2 = \|D_+ A \Delta x_{\rm nt}^+\|^2$ $\leq \|D_{+}A\Delta x_{nt}^{+}\|^{2} + \|D_{+}A\Delta x_{nt}^{+} + (I - D_{+}^{-1}D)DA\Delta x_{nt}\|^{2}$ $= \|(I - D_{\perp}^{-1}D)DA\Delta x_{nt}\|^{2}$

The second inequality follows from $\sum_i y_i^4 \le (\sum_i y_i^2)^2$

bound on $\lambda_t(x^+)$: we use $D := D_X = \operatorname{diag}(d_X)$ and $D_+ := D_{X^+} = \operatorname{diag}(d_{X^+})$ $\lambda_t(x^+)^2 = \|D_+ A \Delta x_{\mathsf{nt}}^+\|^2$ $\leq \|D_+ A \Delta x_{\mathsf{nt}}^+\|^2 + \|D_+ A \Delta x_{\mathsf{nt}}^+ + (I - D_+^{-1}D)DA\Delta x_{\mathsf{nt}}\|^2$ $= \|(I - D_+^{-1}D)DA\Delta x_{\mathsf{nt}}\|^2$ $= \|(I - D_+^{-1}D)^2\vec{1}\|^2$

The second inequality follows from $\sum_{i} y_{i}^{4} \leq (\sum_{i} y_{i}^{2})^{2}$

bound on $\lambda_t(x^+)$: we use $D := D_x = \operatorname{diag}(d_x)$ and $D_+ := D_{x^+} = \operatorname{diag}(d_{x^+})$ $\lambda_t(x^+)^2 = \|D_+ A \Delta x_{\rm nt}^+\|^2$ $\leq \|D_{+}A\Delta x_{nt}^{+}\|^{2} + \|D_{+}A\Delta x_{nt}^{+} + (I - D_{+}^{-1}D)DA\Delta x_{nt}\|^{2}$ $= \|(I - D_{\perp}^{-1}D)DA\Delta x_{nt}\|^{2}$ $= \|(I - D_{\perp}^{-1}D)^2\vec{1}\|^2$ $\leq \|(I - D_{\perp}^{-1}D)\vec{1}\|^4$

The second inequality follows from $\sum_i y_i^4 \leq (\sum_i y_i^2)^2$

bound on $\lambda_t(x^+)$: we use $D := D_x = \operatorname{diag}(d_x)$ and $D_+ := D_{x^+} = \operatorname{diag}(d_{x^+})$

$$\lambda_{t}(x^{+})^{2} = \|D_{+}A\Delta x_{nt}^{+}\|^{2}$$

$$\leq \|D_{+}A\Delta x_{nt}^{+}\|^{2} + \|D_{+}A\Delta x_{nt}^{+} + (I - D_{+}^{-1}D)DA\Delta x_{nt}\|^{2}$$

$$= \|(I - D_{+}^{-1}D)DA\Delta x_{nt}\|^{2}$$

$$= \|(I - D_{+}^{-1}D)^{2}\vec{1}\|^{2}$$

$$\leq \|(I - D_{+}^{-1}D)\vec{1}\|^{4}$$

$$= \|DA\Delta x_{nt}\|^{4}$$

$$= \lambda_{t}(x)^{4}$$

The second inequality follows from $\sum_i y_i^4 \leq (\sum_i y_i^2)^2$

bound on $\lambda_t(x^+)$:

we use $D := D_X = \operatorname{diag}(d_X)$ and $D_+ := D_{X^+} = \operatorname{diag}(d_{X^+})$

$$\lambda_{t}(x^{+})^{2} = \|D_{+}A\Delta x_{nt}^{+}\|^{2}$$

$$\leq \|D_{+}A\Delta x_{nt}^{+}\|^{2} + \|D_{+}A\Delta x_{nt}^{+} + (I - D_{+}^{-1}D)DA\Delta x_{nt}\|^{2}$$

$$= \|(I - D_{+}^{-1}D)DA\Delta x_{nt}\|^{2}$$

$$= \|(I - D_{+}^{-1}D)^{2}\vec{1}\|^{2}$$

$$\leq \|(I - D_{+}^{-1}D)\vec{1}\|^{4}$$

$$= \|DA\Delta x_{nt}\|^{4}$$

$$= \lambda_{t}(x)^{4}$$

The second inequality follows from $\sum_{i} y_{i}^{4} \leq (\sum_{i} y_{i}^{2})^{2}$

If $\lambda_t(x)$ is large we do not have a guarantee.

Try to avoid this case!!!

Path-following Methods

Try to slowly travel along the central path.

Algorithm 1 PathFollowing

1: start at analytic center

2: while solution not good enough do

3: make step to improve objective function

4: recenter to return to central path

Short Step Barrier Method

simplifying assumptions:

- a first central point $x^*(t_0)$ is given
- $\triangleright x^*(t)$ is computed exactly in each iteration

 $\boldsymbol{\epsilon}$ is approximation we are aiming for

start at $t=t_0$, repeat until $m/t \le \epsilon$

- compute $x^*(\mu t)$ using Newton starting from $x^*(t)$
- $ightharpoonup t := \mu t$

where $\mu = 1 + 1/(2\sqrt{m})$

Short Step Barrier Method

gradient of
$$f_{t+}$$
 at $(x = x^*(t))$

$$\nabla f_{t+}(x) = \nabla f_t(x) + (\mu - 1)tc$$
$$= -(\mu - 1)A^T D_x \vec{1}$$

This holds because $0 = \nabla f_t(x) = tc + A^T D_x \vec{1}$.

The Newton decrement is

$$\lambda_{t+}(x)^{2} = \nabla f_{t+}(x)^{T} H^{-1} \nabla f_{t+}(x)$$

$$= (\mu - 1)^{2} \vec{1}^{T} B (B^{T} B)^{-1} B^{T} \vec{1} \qquad B = D_{x}^{T} A^{T} A^{T$$

This means we are in the range of quadratic convergence!!!

gradient of
$$f_{t+}$$
 at $(x = x^*(t))$

$$\nabla f_{t+}(x) = \nabla f_t(x) + (\mu - 1)tc$$
$$= -(\mu - 1)A^T D_x \vec{1}$$

This holds because $0 = \nabla f_t(x) = tc + A^T D_x \vec{1}$.

The Newton decrement is

$$\lambda_{t^{+}}(x)^{2} = \nabla f_{t^{+}}(x)^{T} H^{-1} \nabla f_{t^{+}}(x)$$

$$= (\mu - 1)^{2} \vec{1}^{T} B (B^{T} B)^{-1} B^{T} \vec{1} \qquad B = D_{x}^{T} A^{T}$$

$$\leq (\mu - 1)^{2} m$$

$$= 1/4$$

gradient of
$$f_{t+}$$
 at $(x = x^*(t))$

$$\nabla f_{t+}(x) = \nabla f_t(x) + (\mu - 1)tc$$
$$= -(\mu - 1)A^T D_X \vec{1}$$

This holds because $0 = \nabla f_t(x) = tc + A^T D_x \vec{1}$.

The Newton decrement is

$$\lambda_{t^{+}}(x)^{2} = \nabla f_{t^{+}}(x)^{T} H^{-1} \nabla f_{t^{+}}(x)$$

$$= (\mu - 1)^{2} \vec{1}^{T} B (B^{T} B)^{-1} B^{T} \vec{1} \qquad B = D_{x}^{T} A$$

$$\leq (\mu - 1)^{2} m$$

$$= 1/4$$

gradient of
$$f_{t+}$$
 at $(x = x^*(t))$

$$\nabla f_{t+}(x) = \nabla f_t(x) + (\mu - 1)tc$$
$$= -(\mu - 1)A^T D_x \vec{1}$$

This holds because $0 = \nabla f_t(x) = tc + A^T D_x \vec{1}$.

The Newton decrement is

$$\lambda_{t^{+}}(x)^{2} = \nabla f_{t^{+}}(x)^{T} H^{-1} \nabla f_{t^{+}}(x)$$

$$= (\mu - 1)^{2} \vec{1}^{T} B (B^{T} B)^{-1} B^{T} \vec{1} \qquad B = D_{x}^{T} A$$

$$\leq (\mu - 1)^{2} m$$

$$= 1/4$$

gradient of
$$f_{t+}$$
 at $(x = x^*(t))$

$$\nabla f_{t+}(x) = \nabla f_t(x) + (\mu - 1)tc$$
$$= -(\mu - 1)A^T D_x \vec{1}$$

This holds because $0 = \nabla f_t(x) = tc + A^T D_x \vec{1}$.

The Newton decrement is

$$\lambda_{t^{+}}(x)^{2} = \nabla f_{t^{+}}(x)^{T} H^{-1} \nabla f_{t^{+}}(x)$$

$$= (\mu - 1)^{2} \vec{1}^{T} B (B^{T} B)^{-1} B^{T} \vec{1} \qquad B = D_{x}^{T} A$$

$$\leq (\mu - 1)^{2} m$$

$$= 1/4$$

Number of Iterations

the number of Newton iterations per outer iteration is very small; in practise only 1 or 2

Number of outer iterations:

We need $t_k = \mu^k t_0 \ge m/\epsilon$. This holds when

$$k \ge \frac{\log(m/(\epsilon t_0))}{\log(\mu)}$$

We get a bound of

$$\mathcal{O}\!\left(\sqrt{m}\log\frac{m}{\epsilon t_0}\right)$$

We show how to get a starting point with $t_0=1/2^L$. Together with $\epsilon\approx 2^{-L}$ we get $\mathcal{O}(L\sqrt{m})$ iterations.

For $x \in P^{\circ}$ and direction $v \neq 0$ define

$$\sigma_X(v) := \max_i \frac{a_i^T v}{s_i(x)}$$

Observation:

$$x + \alpha v \in P$$
 for $\alpha \in \{0, 1/\sigma_x(v)\}$

Suppose that we move from x to $x + \alpha v$. The linear estimate says that $f_t(x)$ should change by $\nabla f_t(x)^T \alpha v$.

$$f_t(x + \alpha v) - f_t(x) = tc^T \alpha v + \phi(x + \alpha v) - \phi(x)$$

$$\phi(x + \alpha v) - \phi(x)$$

Suppose that we move from x to $x + \alpha v$. The linear estimate says that $f_t(x)$ should change by $\nabla f_t(x)^T \alpha v$.

$$f_t(x + \alpha v) - f_t(x) = tc^T \alpha v + \phi(x + \alpha v) - \phi(x)$$

$$\phi(x + \alpha v) - \phi(x)$$

Suppose that we move from x to $x + \alpha v$. The linear estimate says that $f_t(x)$ should change by $\nabla f_t(x)^T \alpha v$.

$$f_t(x + \alpha v) - f_t(x) = tc^T \alpha v + \phi(x + \alpha v) - \phi(x)$$

$$\phi(x + \alpha v) - \phi(x) = -\sum_i \log(s_i(x + \alpha v)) + \sum_i \log(s_i(x))$$

$$= -\sum_i \log(s_i(x + \alpha v)/s_i(x))$$

$$= -\sum_i \log(1 - \alpha_i^T \alpha v/s_i(x))$$

Suppose that we move from x to $x + \alpha v$. The linear estimate says that $f_t(x)$ should change by $\nabla f_t(x)^T \alpha v$.

$$f_t(x + \alpha v) - f_t(x) = tc^T \alpha v + \phi(x + \alpha v) - \phi(x)$$

$$\phi(x + \alpha v) - \phi(x) = -\sum_i \log(s_i(x + \alpha v)) + \sum_i \log(s_i(x))$$

$$= -\sum_i \log(s_i(x + \alpha v)/s_i(x))$$

$$= -\sum_i \log(1 - \alpha_i^T \alpha v/s_i(x))$$

Suppose that we move from x to $x + \alpha v$. The linear estimate says that $f_t(x)$ should change by $\nabla f_t(x)^T \alpha v$.

$$f_t(x + \alpha v) - f_t(x) = tc^T \alpha v + \phi(x + \alpha v) - \phi(x)$$

$$\phi(x + \alpha v) - \phi(x) = -\sum_i \log(s_i(x + \alpha v)) + \sum_i \log(s_i(x))$$

$$= -\sum_i \log(s_i(x + \alpha v)/s_i(x))$$

$$= -\sum_i \log(1 - a_i^T \alpha v/s_i(x))$$

Define
$$w_i = a_i^T v / s_i(x)$$
 and $\sigma = \max_i w_i$. Then

$$f_t(x + \alpha v) - f_t(x) - \nabla f_t(x)^T \alpha v$$

Define $w_i = a_i^T v / s_i(x)$ and $\sigma = \max_i w_i$. Then

$$\begin{split} f_t(x + \alpha v) - f_t(x) - \nabla f_t(x)^T \alpha v \\ &= -\sum_i (\alpha w_i + \log(1 - \alpha w_i)) \\ &\leq -\sum_{w_i > 0} (\alpha w_i + \log(1 - \alpha w_i)) + \sum_{w_i \leq 0} \frac{\alpha^2 w_i^2}{2} \\ &\leq -\sum_{w_i > 0} \frac{w_i^2}{\sigma^2} \left(\alpha \sigma + \log(1 - \alpha \sigma)\right) + \frac{(\alpha \sigma)^2}{2} \sum_{w_i \leq 0} \frac{w_i^2}{\sigma^2} \end{split}$$

Define $w_i = a_i^T v / s_i(x)$ and $\sigma = \max_i w_i$. Then

$$\begin{aligned} f_t(x + \alpha v) - f_t(x) - \nabla f_t(x)^T \alpha v \\ &= -\sum_i (\alpha w_i + \log(1 - \alpha w_i)) \\ &\leq -\sum_{w_i > 0} (\alpha w_i + \log(1 - \alpha w_i)) + \sum_{w_i \leq 0} \frac{\alpha^2 w_i^2}{2} \\ &\leq -\sum_{w_i > 0} \frac{w_i^2}{\sigma^2} \left(\alpha \sigma + \log(1 - \alpha \sigma)\right) + \frac{(\alpha \sigma)^2}{2} \sum_{w_i \leq 0} \frac{w_i^2}{\sigma^2} \end{aligned}$$

Define $w_i = a_i^T v / s_i(x)$ and $\sigma = \max_i w_i$. Then

$$\begin{split} f_t(x + \alpha v) - f_t(x) - \nabla f_t(x)^T \alpha v \\ &= -\sum_i (\alpha w_i + \log(1 - \alpha w_i)) \\ &\leq -\sum_{w_i > 0} (\alpha w_i + \log(1 - \alpha w_i)) + \sum_{w_i \leq 0} \frac{\alpha^2 w_i^2}{2} \\ &\leq -\sum_{w_i > 0} \frac{w_i^2}{\sigma^2} \Big(\alpha \sigma + \log(1 - \alpha \sigma)\Big) + \frac{(\alpha \sigma)^2}{2} \sum_{w_i \leq 0} \frac{w_i^2}{\sigma^2} \end{split}$$

$$\leq -\sum_{i} \frac{w_{i}^{2}}{\sigma^{2}} \left(\alpha \sigma + \log(1 - \alpha \sigma) \right)$$

$$= -\frac{1}{\sigma^{2}} \|v\|_{H_{x}}^{2} \left(\alpha \sigma + \log(1 - \alpha \sigma) \right)$$

Damped Newton Iteration: In a damped Newton step we choose

$$x_{+} = x + \frac{1}{1 + \sigma_{X}(\Delta x_{nt})} \Delta x_{nt}$$

$$\leq -\sum_{i} \frac{w_{i}^{2}}{\sigma^{2}} \left(\alpha \sigma + \log(1 - \alpha \sigma) \right)$$
$$= -\frac{1}{\sigma^{2}} \|v\|_{H_{x}}^{2} \left(\alpha \sigma + \log(1 - \alpha \sigma) \right)$$

Damped Newton Iteration: In a damped Newton step we choose

$$x_{+} = x + \frac{1}{1 + \sigma_{x}(\Delta x_{\text{nt}})} \Delta x_{\text{nt}}$$

$$\leq -\sum_{i} \frac{w_{i}^{2}}{\sigma^{2}} \left(\alpha \sigma + \log(1 - \alpha \sigma) \right)$$
$$= -\frac{1}{\sigma^{2}} \|v\|_{H_{x}}^{2} \left(\alpha \sigma + \log(1 - \alpha \sigma) \right)$$

Damped Newton Iteration:

In a damped Newton step we choose

$$x_{+} = x + \frac{1}{1 + \sigma_{x}(\Delta x_{\mathsf{nt}})} \Delta x_{\mathsf{nt}}$$

Theorem:

In a damped Newton step the cost decreases by at least

$$\lambda_t(x) - \log(1 + \lambda_t(x))$$

Proof: The decrease in cost is

$$-\alpha \nabla f_t(x)^T v + \frac{1}{\sigma^2} \|v\|_{H_x}^2 (\alpha \sigma + \log(1 - \alpha \sigma))$$

Choosing $lpha=rac{1}{1+\sigma}$ and $v=\Delta x_{\sf nt}$ gives

Theorem:

In a damped Newton step the cost decreases by at least

$$\lambda_t(x) - \log(1 + \lambda_t(x))$$

Proof: The decrease in cost is

$$-\alpha \nabla f_t(x)^T v + \frac{1}{\sigma^2} \|v\|_{H_x}^2 (\alpha \sigma + \log(1 - \alpha \sigma))$$

Choosing $\alpha = \frac{1}{1+\sigma}$ and $v = \Delta x_{\sf nt}$ gives

Theorem:

In a damped Newton step the cost decreases by at least

$$\lambda_t(x) - \log(1 + \lambda_t(x))$$

Proof: The decrease in cost is

$$-\alpha \nabla f_t(x)^T v + \frac{1}{\sigma^2} \|v\|_{H_x}^2 (\alpha \sigma + \log(1 - \alpha \sigma))$$

Choosing $\alpha = \frac{1}{1+\sigma}$ and $v = \Delta x_{nt}$ gives

$$\frac{1}{1+\sigma}\lambda_t(x)^2 + \frac{\lambda_t(x)^2}{\sigma^2} \left(\frac{\sigma}{1+\sigma} + \log\left(1 - \frac{\sigma}{1+\sigma}\right) \right)$$
$$= \frac{\lambda_t(x)^2}{\sigma^2} \left(\sigma - \log(1+\sigma) \right)$$

Theorem:

In a damped Newton step the cost decreases by at least

$$\lambda_t(x) - \log(1 + \lambda_t(x))$$

Proof: The decrease in cost is

$$-\alpha \nabla f_t(x)^T v + \frac{1}{\sigma^2} \|v\|_{H_x}^2 (\alpha \sigma + \log(1 - \alpha \sigma))$$

Choosing $\alpha = \frac{1}{1+\sigma}$ and $v = \Delta x_{nt}$ gives

$$\frac{1}{1+\sigma}\lambda_t(x)^2 + \frac{\lambda_t(x)^2}{\sigma^2} \left(\frac{\sigma}{1+\sigma} + \log\left(1 - \frac{\sigma}{1+\sigma}\right) \right)$$
$$= \frac{\lambda_t(x)^2}{\sigma^2} \left(\sigma - \log(1+\sigma) \right)$$

$$\geq \lambda_t(x) - \log(1 + \lambda_t(x))$$

$$\geq 0.09$$

for
$$\lambda_t(x) \ge 0.5$$

Centering Algorithm:

Input: precision δ ; starting point x

- **1.** compute $\Delta x_{\rm nt}$ and $\lambda_t(x)$
- **2.** if $\lambda_t(x) \leq \delta$ return x
- 3. set $x := x + \alpha \Delta x_{nt}$ with

$$\alpha = \begin{cases} \frac{1}{1 + \sigma_X(\Delta x_{\mathsf{nt}})} & \lambda_t \ge 1/2 \\ 1 & \mathsf{otw.} \end{cases}$$

$$\geq \lambda_t(x) - \log(1 + \lambda_t(x))$$

$$\geq 0.09$$

for $\lambda_t(x) \ge 0.5$

Centering Algorithm:

Input: precision δ ; starting point x

- **1.** compute Δx_{nt} and $\lambda_t(x)$
- **2.** if $\lambda_t(x) \leq \delta$ return x
- 3. set $x := x + \alpha \Delta x_{nt}$ with

$$\alpha = \left\{ \begin{array}{ll} \frac{1}{1+\sigma_X(\Delta x_{\mathsf{nt}})} & \lambda_t \geq 1/2 \\ 1 & \mathsf{otw.} \end{array} \right.$$

Centering

Lemma 56

The centering algorithm starting at x_0 reaches a point with $\lambda_t(x) \leq \delta$ after

$$\frac{f_t(x_0) - \min_{\mathcal{Y}} f_t(\mathcal{Y})}{0.09} + \mathcal{O}(\log\log(1/\delta))$$

iterations.

This can be very, very slow...

Let $P = \{Ax \le b\}$ be our (feasible) polyhedron, and x_0 a feasible point.

We change $b \to b + \frac{1}{\lambda} \cdot \vec{1}$, where $L = \langle A \rangle + \langle b \rangle + \langle c \rangle$ (encoding length) and $\lambda = 2^{2L}$. Recall that a basis is feasible in the old LP iff it is feasible in the new LP.

Let $P = \{Ax \le b\}$ be our (feasible) polyhedron, and x_0 a feasible point.

We change $b \to b + \frac{1}{\lambda} \cdot \vec{1}$, where $L = \langle A \rangle + \langle b \rangle + \langle c \rangle$ (encoding length) and $\lambda = 2^{2L}$. Recall that a basis is feasible in the old LP iff it is feasible in the new LP.

The inverse of a matrix M can be represented with rational numbers that have denominators $z_{ij} = \det(M)$.

For two basis solutions x_B , $x_{\bar{B}}$, the cost-difference $c^Tx_B - c^Tx_{\bar{B}}$ can be represented by a rational number that has denominator $z = \det(A_B) \cdot \det(A_{\bar{B}})$.

This means that in the perturbed LP it is sufficient to decrease the duality gap to $1/2^{4L}$ (i.e., $t\approx 2^{4L}$). This means the previous analysis essentially also works for the perturbed LP.

The inverse of a matrix M can be represented with rational numbers that have denominators $z_{ij} = \det(M)$.

For two basis solutions x_B , $x_{\bar{B}}$, the cost-difference $c^Tx_B - c^Tx_{\bar{B}}$ can be represented by a rational number that has denominator $z = \det(A_B) \cdot \det(A_{\bar{B}})$.

This means that in the perturbed LP it is sufficient to decrease the duality gap to $1/2^{4L}$ (i.e., $t\approx 2^{4L}$). This means the previous analysis essentially also works for the perturbed LP.

The inverse of a matrix M can be represented with rational numbers that have denominators $z_{ij} = \det(M)$.

For two basis solutions x_B , $x_{\bar{B}}$, the cost-difference $c^Tx_B - c^Tx_{\bar{B}}$ can be represented by a rational number that has denominator $z = \det(A_B) \cdot \det(A_{\bar{B}})$.

This means that in the perturbed LP it is sufficient to decrease the duality gap to $1/2^{4L}$ (i.e., $t\approx 2^{4L}$). This means the previous analysis essentially also works for the perturbed LP.

The inverse of a matrix M can be represented with rational numbers that have denominators $z_{ij} = \det(M)$.

For two basis solutions x_B , $x_{\bar{B}}$, the cost-difference $c^Tx_B - c^Tx_{\bar{B}}$ can be represented by a rational number that has denominator $z = \det(A_B) \cdot \det(A_{\bar{B}})$.

This means that in the perturbed LP it is sufficient to decrease the duality gap to $1/2^{4L}$ (i.e., $t\approx 2^{4L}$). This means the previous analysis essentially also works for the perturbed LP.

Start at x_0 .

Choose
$$\hat{c} := -\nabla \phi(x)$$
.

$$x_0 = x^*(1)$$
 is point on central path for \hat{c} and $t = 1$.

You can travel the central path in both directions. Go towards 0 until $t \approx 1/2^{\Omega(L)}$. This requires $O(\sqrt{m}L)$ outer iterations.

Let $x_{\hat{c}}$ denote this point.

Let x_c denote the point that minimizes

$$t \cdot c^T x + \phi(x)$$

Start at x_0 .

Choose $\hat{c} := -\nabla \phi(x)$.

 $x_0 = x^*(1)$ is point on central path for \hat{c} and t = 1.

You can travel the central path in both directions. Go towards 0 until $t \approx 1/2^{\Omega(L)}$. This requires $O(\sqrt{m}L)$ outer iterations.

Let $x_{\hat{c}}$ denote this point.

Let x_c denote the point that minimizes

$$t \cdot c^T x + \phi(x)$$

Start at x_0 .

Choose
$$\hat{c} := -\nabla \phi(x)$$
.

$$x_0 = x^*(1)$$
 is point on central path for \hat{c} and $t = 1$.

You can travel the central path in both directions. Go towards 0 until $t \approx 1/2^{\Omega(L)}$. This requires $O(\sqrt{m}L)$ outer iterations.

Let $x_{\hat{c}}$ denote this point

Let x_c denote the point that minimizes

$$t \cdot c^T x + \phi(x)$$

Start at x_0 .

Choose
$$\hat{c} := -\nabla \phi(x)$$
.

$$x_0 = x^*(1)$$
 is point on central path for \hat{c} and $t = 1$.

You can travel the central path in both directions. Go towards 0 until $t \approx 1/2^{\Omega(L)}$. This requires $O(\sqrt{m}L)$ outer iterations.

Let $x_{\hat{c}}$ denote this point.

Let x_c denote the point that minimizes

$$t \cdot c^T x + \phi(x)$$

Start at x_0 .

Choose $\hat{c} := -\nabla \phi(x)$.

 $x_0 = x^*(1)$ is point on central path for \hat{c} and t = 1.

You can travel the central path in both directions. Go towards 0 until $t \approx 1/2^{\Omega(L)}$. This requires $O(\sqrt{m}L)$ outer iterations.

Let $x_{\hat{c}}$ denote this point.

Let x_c denote the point that minimizes

$$t \cdot c^T x + \phi(x)$$

Start at x_0 .

Choose $\hat{c} := -\nabla \phi(x)$.

 $x_0 = x^*(1)$ is point on central path for \hat{c} and t = 1.

You can travel the central path in both directions. Go towards 0 until $t \approx 1/2^{\Omega(L)}$. This requires $O(\sqrt{m}L)$ outer iterations.

Let $x_{\hat{c}}$ denote this point.

Let x_c denote the point that minimizes

$$t \cdot c^T x + \phi(x)$$

Clearly,

$$t \cdot \hat{c}^T x_{\hat{c}} + \phi(x_{\hat{c}}) \leq t \cdot \hat{c}^T x_c + \phi(x_c)$$

The difference between $f_t(x_{\hat{c}})$ and $f_t(x_c)$ is

$$tc^{T}x_{\hat{c}} + \phi(x_{\hat{c}}) - tc^{T}x_{c} - \phi(x_{c})$$

$$\leq t(c^{T}x_{\hat{c}} + \hat{c}^{T}x_{c} - \hat{c}^{T}x_{\hat{c}} - c^{T}x_{c})$$

$$\leq 4tn2^{3L}$$

For $t=1/2^{\Omega(L)}$ the last term becomes constant. Hence, using damped Newton we can move from $x_{\hat{c}}$ to $x_{\hat{c}}$ quickly.

In total for this analysis we require $\mathcal{O}(\sqrt{m}L)$ outer iterations for the whole algorithm.

One iteration can be implemented in $\tilde{\mathcal{O}}(m^3)$ time

Clearly,

$$t \cdot \hat{c}^T x_{\hat{c}} + \phi(x_{\hat{c}}) \leq t \cdot \hat{c}^T x_c + \phi(x_c)$$

The difference between $f_t(x_{\hat{c}})$ and $f_t(x_c)$ is

$$tc^{T}x_{\hat{c}} + \phi(x_{\hat{c}}) - tc^{T}x_{c} - \phi(x_{c})$$

$$\leq t(c^{T}x_{\hat{c}} + \hat{c}^{T}x_{c} - \hat{c}^{T}x_{\hat{c}} - c^{T}x_{c})$$

$$\leq 4tn2^{3L}$$

For $t=1/2^{\Omega(L)}$ the last term becomes constant. Hence, using damped Newton we can move from $x_{\hat{c}}$ to $x_{\hat{c}}$ quickly.

In total for this analysis we require $\mathcal{O}(\sqrt{m}L)$ outer iterations for the whole algorithm.

One iteration can be implemented in $\tilde{\mathcal{O}}(m^3)$ time

Clearly,

$$t \cdot \hat{c}^T x_{\hat{c}} + \phi(x_{\hat{c}}) \le t \cdot \hat{c}^T x_c + \phi(x_c)$$

The difference between $f_t(x_{\hat{c}})$ and $f_t(x_c)$ is

$$tc^{T}x_{\hat{c}} + \phi(x_{\hat{c}}) - tc^{T}x_{c} - \phi(x_{c})$$

$$\leq t(c^{T}x_{\hat{c}} + \hat{c}^{T}x_{c} - \hat{c}^{T}x_{\hat{c}} - c^{T}x_{c})$$

$$\leq 4tn2^{3L}$$

For $t=1/2^{\Omega(L)}$ the last term becomes constant. Hence, using damped Newton we can move from $x_{\hat{c}}$ to $x_{\hat{c}}$ quickly.

In total for this analysis we require $\mathcal{O}(\sqrt{m}L)$ outer iterations for the whole algorithm.

One iteration can be implemented in $\tilde{\mathcal{O}}(m^3)$ time.

Clearly,

$$t \cdot \hat{c}^T x_{\hat{c}} + \phi(x_{\hat{c}}) \le t \cdot \hat{c}^T x_c + \phi(x_c)$$

The difference between $f_t(x_{\hat{c}})$ and $f_t(x_c)$ is

$$tc^{T}x_{\hat{c}} + \phi(x_{\hat{c}}) - tc^{T}x_{c} - \phi(x_{c})$$

$$\leq t(c^{T}x_{\hat{c}} + \hat{c}^{T}x_{c} - \hat{c}^{T}x_{\hat{c}} - c^{T}x_{c})$$

$$\leq 4tn2^{3L}$$

For $t=1/2^{\Omega(L)}$ the last term becomes constant. Hence, using damped Newton we can move from $x_{\hat{c}}$ to $x_{\hat{c}}$ quickly.

In total for this analysis we require $\mathcal{O}(\sqrt{m}L)$ outer iterations for the whole algorithm.

One iteration can be implemented in $\tilde{\mathcal{O}}(m^3)$ time

Clearly,

$$t \cdot \hat{c}^T x_{\hat{c}} + \phi(x_{\hat{c}}) \leq t \cdot \hat{c}^T x_c + \phi(x_c)$$

The difference between $f_t(x_{\hat{c}})$ and $f_t(x_c)$ is

$$tc^{T}x_{\hat{c}} + \phi(x_{\hat{c}}) - tc^{T}x_{c} - \phi(x_{c})$$

$$\leq t(c^{T}x_{\hat{c}} + \hat{c}^{T}x_{c} - \hat{c}^{T}x_{\hat{c}} - c^{T}x_{c})$$

$$\leq 4tn2^{3L}$$

For $t=1/2^{\Omega(L)}$ the last term becomes constant. Hence, using damped Newton we can move from $x_{\hat{c}}$ to $x_{\mathcal{C}}$ quickly.

In total for this analysis we require $\mathcal{O}(\sqrt{m}L)$ outer iterations for the whole algorithm.

One iteration can be implemented in $\tilde{\mathcal{O}}(m^3)$ time.

Clearly,

$$t \cdot \hat{c}^T x_{\hat{c}} + \phi(x_{\hat{c}}) \le t \cdot \hat{c}^T x_c + \phi(x_c)$$

The difference between $f_t(x_{\hat{c}})$ and $f_t(x_c)$ is

$$tc^{T}x_{\hat{c}} + \phi(x_{\hat{c}}) - tc^{T}x_{c} - \phi(x_{c})$$

$$\leq t(c^{T}x_{\hat{c}} + \hat{c}^{T}x_{c} - \hat{c}^{T}x_{\hat{c}} - c^{T}x_{c})$$

$$\leq 4tn2^{3L}$$

For $t=1/2^{\Omega(L)}$ the last term becomes constant. Hence, using damped Newton we can move from $x_{\hat{c}}$ to x_{c} quickly.

In total for this analysis we require $\mathcal{O}(\sqrt{m}L)$ outer iterations for the whole algorithm.

One iteration can be implemented in $\tilde{\mathcal{O}}(m^3)$ time

Clearly,

$$t \cdot \hat{c}^T x_{\hat{c}} + \phi(x_{\hat{c}}) \le t \cdot \hat{c}^T x_c + \phi(x_c)$$

The difference between $f_t(x_{\hat{c}})$ and $f_t(x_c)$ is

$$tc^{T}x_{\hat{c}} + \phi(x_{\hat{c}}) - tc^{T}x_{c} - \phi(x_{c})$$

$$\leq t(c^{T}x_{\hat{c}} + \hat{c}^{T}x_{c} - \hat{c}^{T}x_{\hat{c}} - c^{T}x_{c})$$

$$\leq 4tn2^{3L}$$

For $t=1/2^{\Omega(L)}$ the last term becomes constant. Hence, using damped Newton we can move from $x_{\hat{c}}$ to $x_{\hat{c}}$ quickly.

In total for this analysis we require $\mathcal{O}(\sqrt{m}L)$ outer iterations for the whole algorithm.

One iteration can be implemented in $\tilde{\mathcal{O}}(m^3)$ time.