Machine Learning II

Bjoern Andres

Machine Learning for Computer Vision
TU Dresden

https://mlcv.cs.tu-dresden.de/courses/26-summer/ml2/

Summer Term 2026

Definition 1. For any $A \in \mathbb{R}^{m \times n}$ and any $b \in \mathbb{R}^m$,

$$A\,x = b \tag{1}$$

$$x \geq 0 \tag{2}$$

is called the **standard system** defined by A and b, and

$$X_{Ab} := \{x \in \mathbb{R}^n \mid A\,x = b \wedge x \geq 0\} \tag{3}$$

is called its **feasible set**. The elements of X_{Ab} are called the **feasible solutions** to (1)–(2).

Remark 1. If the rows of A are linearly dependent, rows can be deleted without changing feasible set X_{Ab}. Thus, we may assume that rank$(A) = m$.

Remark 2. Assume rank$(A) = m$. If $n \leq m$ then $n = m$. In this case, the solution $x' := A^{-1}b$ to (1) is unique. If $x' \geq 0$, $X_{Ab} = \{x'\}$. Otherwise, $X_{Ab} = \varnothing$. More interesting is the case where $n > m$.

Definition 2. For any $m, n \in \mathbb{N}_0$ and any $A \in \mathbb{R}^{m \times n}$ with $\text{rank}(A) = m$, a **basis** of A is any $K \subseteq n$ with $|K| = m$ such that $A|_{\cdot K}$ is non-singular.

Definition 3. For any $m, n \in \mathbb{N}_0$, any $A \in \mathbb{R}^{m \times n}$ with $\text{rank}(A) = m$, any $b \in \mathbb{R}^m$, any $x \in X_{Ab}$ and any basis K of A, x is called a **basic feasible solution (BFS)** with basis K to the standard system defined by A and b iff

$$x|_K = A|_{\cdot K}^{-1} b \tag{4}$$

$$x|_{n \setminus K} = 0 \tag{5}$$

Lemma 1. For any $m, n \in \mathbb{N}_0$, any $A \in \mathbb{R}^{m \times n}$ with $\text{rank}(A) = m$ and any $b \in \mathbb{R}^m$, the standard system defined by A and b has at most $\binom{n}{m}$ BFSs.

Proof. Every BFS is related to a basis.

For any basis K, the $x \in \mathbb{R}^n$ defined by (4)–(5) is unique. x is a BFS iff $x \in X_{Ab}$. Thus, there is at most one BFS per basis.

Every basis K is, in particular, a set $K \subseteq n$ with $|K| = m$, of which there are $\binom{n}{m}$ many. □

Theorem 1. Let $m, n \in \mathbb{N}_0$, $A \in \mathbb{R}^{m \times n}$ with $\operatorname{rank}(A) = m$ and $b \in \mathbb{R}^m$. If $X_{Ab} \neq \varnothing$, there exists a BFS $x \in X_{Ab}$.

Remark 3. By virtue of Theorem 1 and Lemma 1, the search for a feasible solution can be reduced to a finite set.

Proof. For any $y \in \mathbb{R}^n$, let $K_y := \{k \in n \mid y_k \neq 0\}$. Consider any $x \in X_{Ab}$.

a) If the columns of $A|_{\cdot K_x}$ are l.i., rank$(A) = m$ implies $|K_x| \leq m$.

- If $|K_x| = m$, x is a BFS with the basis K_x.
- If $|K_x| < m$, choose $K \subseteq n$ such that $K_x \subseteq K$ and $|K| = m$ such that the columns of $A|_{\cdot K}$ are l.i.. This is possible because rank$(A) = m$. Now, x is a BFS with the basis K.

b) If the columns of $A|_{\cdot K_x}$ are l.d., there exists $d \in \mathbb{R}^n$ such that

$$A\, d = 0 \tag{6}$$

$$\varnothing \neq K_d \subseteq K_x \ . \tag{7}$$

For any $\lambda \in \mathbb{R}$, let $x' := x + \lambda\, d$. Now:

$$A\, x' = A\, x + \lambda\, A\, d \overset{(6)}{=} A\, x = b \ . \tag{8}$$

Choose $k' \in K_d$ so as to minimize $\frac{x_{k'}}{|d_{k'}|}$. Let $\lambda' := -\frac{x_{k'}}{d_{k'}}$. Now:

$$\forall k \in m \setminus K_x: \quad x'_k = x_k - \lambda' d_k \overset{(7)}{=} 0 \tag{9}$$

$$x'_{k'} = x_{k'} - \frac{x_{k'}}{d_{k'}} d_{k'} = 0 \tag{10}$$

$$\forall k \in K_x \setminus \{k'\}: \quad x'_k = x_k - \frac{x_{k'}}{d_{k'}} d_k = \begin{cases} x_k & \text{if } d_k = 0 \\ \left(\frac{x_k}{d_k} - \frac{x_{k'}}{d_{k'}}\right) d_k & \text{if } d_k \neq 0 \ . \end{cases} \tag{11}$$

Proof (contd.). We analyze (11): If $d_k > 0$ then

$$\left(\frac{x_k}{d_k} - \frac{x_{k'}}{d_{k'}}\right) d_k \;=\; \left(\frac{x_k}{|d_k|} - \frac{x_{k'}}{d_{k'}}\right) |d_k| \;\geq\; \left(\frac{x_k}{|d_k|} - \frac{x_{k'}}{|d_{k'}|}\right) |d_k| \;\geq\; 0 \;. \qquad (12)$$

If $d_k < 0$ then

$$\begin{aligned}\left(\frac{x_k}{d_k} - \frac{x_{k'}}{d_{k'}}\right) d_k \;&=\; \left(-\frac{x_k}{d_k} + \frac{x_{k'}}{d_{k'}}\right) |d_k| \;=\; \left(\frac{x_k}{|d_k|} + \frac{x_{k'}}{d_{k'}}\right) |d_k| \\ &\geq\; \left(\frac{x_k}{|d_k|} - \frac{x_{k'}}{|d_{k'}|}\right) |d_k| \;\geq\; 0 \;. \qquad (13)\end{aligned}$$

Thus, $x' \geq 0$. Together with (8) follows $x' \in X_{Ab}$. Moreover, $K_{x'} \subseteq K_x \setminus \{k'\}$.

If the columns of $A|_{\cdot K_{x'}}$ are l.i., a BFS exists by Part (a).

Otherwise, Part (b) can be iterated until the columns of $A|_{\cdot K_{x'}}$ are l.i..

Thus, a BFS always exists. □

Definition 4. Let $m, n \in \mathbb{N}_0$, $A \in \mathbb{R}^{m \times n}$ with $\text{rank}(A) = m$ and $b \in \mathbb{R}^m$. For

- any BFS x with basis K to the standard system defined by A and b
- any $k^* \in n$ and the unique $y^{k^*} \in \mathbb{R}^n$ such that

$$\sum_{k \in K} y_k^{k^*} A_{\cdot k} = A_{\cdot k^*} \qquad \text{i.e. } y^{k^*}|_K = A|_{\cdot K}^{-1} A_{\cdot k^*} \tag{14}$$

$$\forall k \in n \setminus K\colon \qquad y_k^{k^*} = 0 \tag{15}$$

- any $k^\dagger \in K_{k^*} := \{k \in K \mid y_k^{k^*} > 0\}$ and the unique $\lambda' \in \mathbb{R}_0^+$ such that

$$\min \left\{ \frac{x_k}{y_k^{k^*}} \;\middle|\; k \in K_{k^*} \right\} = \frac{x_{k^\dagger}}{y_{k^\dagger}^{k^*}} = \lambda' \;, \tag{16}$$

the $x' \in \mathbb{R}^n$ such that

$$x'_{k^*} := \lambda' \tag{17}$$

$$\forall k \in n \setminus \{k^*\}\colon \quad x'_k := x_k - \lambda' y_k^{k^*} \tag{18}$$

is called the result of **pivoting** x in K such that k^* **enters** and $k^\dagger$ **leaves** the basis.

Theorem 2. Let

- $m, n \in \mathbb{N}_0$, $A \in \mathbb{R}^{m \times n}$ with rank$(A) = m$ and $b \in \mathbb{R}^m$
- x a BFS with basis K to the standard system defined by A and b
- x' be the result of pivoting x in K such that k^* enters and $k^\dagger$ leaves the basis.

Then, x' is a BFS with basis $(K \setminus \{k^\dagger\}) \cup \{k^*\}$ to the standard system defined by A and b.

Remark 4. Consider Definition 4.

a) If $K_{k^*} = \varnothing$ then, in (17)–(18), $x' \geq 0$ for all $\lambda' \in \mathbb{R}_0^+$. Thus, X_{Ab} is unbounded.

b) If $\lambda' = 0$ then $x' = x$ (while k^* enters and $k^\dagger$ leaves the basis).

c) If $k^* \in K$ then

$$\forall k \in K\colon \quad y_k^{k^*} = \begin{cases} 1 & \text{if } k = k^* \\ 0 & \text{otherwise} \end{cases} . \tag{19}$$

Therefore, $K_{k^*} = \{k^*\}$. Thus, $k^\dagger = k^*$ and $\lambda' = x_{k^*}$. Hence, the basis does not change and $x' = x$.

Proof. Since x is a BFS with basis K:

$$\sum_{k \in K} x_k \, A_{\cdot k} = b \ . \tag{20}$$

For any $\lambda \in \mathbb{R}$, (20)–λ(14) yields

$$\lambda A_{\cdot k^*} + \sum_{k \in K} (x_k - \lambda y_k^{k^*}) \, A_{\cdot k} = b \ . \tag{21}$$

For $\lambda = \lambda'$ specifically follows $A \, x' = b$.

From $x_{k^\dagger} \geq 0$ and $y_{k^\dagger}^{k^*} > 0$ follows $\lambda' \geq 0$. Thus $x'_{k^\dagger} \geq 0$. For any $k \in n \setminus \{k^\dagger\}$, distinguish two cases: If $y_k^{k^*} \leq 0$ then (18), $x_k \geq 0$ and $\lambda' \geq 0$ imply $x'_k \geq 0$. If $y_k^{k^*} > 0$ then

$$x'_k = x_k - \lambda' y_k^{k^*} = x_k - \frac{x_{k^\dagger}}{y_{k^\dagger}^{k^*}} y_k^{k^*} = \left(\frac{x_k}{y_k^{k^*}} - \frac{x_{k^\dagger}}{y_{k^\dagger}^{k^*}} \right) y_k^{k^*} \overset{(16)}{\geq} 0 \ . \tag{22}$$

Thus, $x' \geq 0$.

It remains to be shown that $(K \setminus \{k^\dagger\}) \cup \{k^*\}$ is a basis. ...

Proof (contd.). Let $c \in \mathbb{R}^n$ such that

$$0 = \sum_{k \in (K \setminus \{k^\dagger\}) \cup \{k^*\}} c_k \, A_{\cdot k} \quad . \tag{23}$$

Now,

$$0 = c_{k^*} A_{\cdot k^*} + \sum_{k \in K \setminus \{k^\dagger\}} c_k A_{\cdot k} \tag{24}$$

$$\overset{(14)}{=} c_{k^*} \sum_{k \in K} y_k^{k^*} A_{\cdot k} + \sum_{k \in K \setminus \{k^\dagger\}} c_k A_{\cdot k} \tag{25}$$

$$= c_{k^*} y_{k^\dagger}^{k^*} A_{\cdot k^\dagger} + \sum_{k \in K \setminus \{k^\dagger\}} (c_{k^*} y_k^{k^*} + c_k) A_{\cdot k} \quad . \tag{26}$$

The rhs. is a l.c. of the columns of A indexed by the basis K. Thus,

$$c_{k^*} y_{k^\dagger}^{k^*} = 0 \tag{27}$$

$$\forall k \in K \setminus \{k^\dagger\}\colon \quad c_{k^*} y_k^{k^*} + c_k = 0 \quad . \tag{28}$$

From (27) and $y_{k^\dagger}^{k^*} > 0$ follows $c_{k^*} = 0$. Together with (28) follows $c = 0$. Thus, $(K \setminus \{k^\dagger\}) \cup \{k^*\}$ is a basis. □

Definition 5. For any $m, n \in \mathbb{N}_0$, any $A \in \mathbb{R}^{m \times n}$, any $A' \in \mathbb{R}^{m' \times n}$, any $b \in \mathbb{R}^m$, any $b' \in \mathbb{R}^{m'}$ and any $J \subseteq n$, the **general system** defined by A, A', b, b' and J is

$$A\,x = b \tag{29}$$
$$A'\,x \geq b' \tag{30}$$
$$\forall j \in J: \quad x_j \geq 0 \ . \tag{31}$$

Definition 6. For any $A \in \mathbb{R}^{m \times n}$ and any $b \in \mathbb{R}^m$, the **canonical system** defined by A and b is

$$A\,x \geq b \tag{32}$$
$$x \geq 0 \ . \tag{33}$$

Lemma 2. Standard, general, and canonical systems are mutually equivalent, in the sense that each can be transformed into any of the others.

Proof. A canonical system is a special case of a general system.

To turn any general system into an equivalent canonical system, proceed as follows:

- For any $j \in n \setminus J$, replace the unconstrained variable x_j by $x'_j - x''_j$ such that

$$x'_j \geq 0 \tag{34}$$
$$x''_j \geq 0 \ . \tag{35}$$

- Replace any equality $A_{j\cdot}\, x = b_j$ by the inequalities

$$A_{j\cdot}\, x \geq b_j \tag{36}$$
$$-A_{j\cdot}\, x \geq -b_j \ . \tag{37}$$

A standard system is a special case of a general system.

To turn any general system into an equivalent standard system, proceed as follows:

- Replace all unconstrained variables as described above.
- Replace any inequality $A_{j\cdot}\, x \geq b_j$ by

$$A_{j\cdot}\, x - x'''_j = b_j \tag{38}$$
$$x'''_j \geq 0 \ . \tag{39}$$

□

Definition 7. For any $m, n \in \mathbb{N}_0$, any $A \in \mathbb{R}^{m \times n}$, any $b \in \mathbb{R}^m$ and any $c \in \mathbb{R}^n$, the **linear optimization problem** or **linear program (LP)** in **standard form** is

$$\min \left\{ \langle c, x \rangle \;\middle|\; A\,x = b \;\wedge\; x \geq 0 \;\wedge\; x \in \mathbb{R}^n \right\} \;, \tag{40}$$

and the linear optimization problem or linear program (LP) in **canonical form** is

$$\min \left\{ \langle c, x \rangle \;\middle|\; A\,x \geq b \;\wedge\; x \geq 0 \;\wedge\; x \in \mathbb{R}^n \right\} \;. \tag{41}$$

For any $m, n \in \mathbb{N}_0$, any $A \in \mathbb{R}^{m \times n}$, any $A' \in \mathbb{R}^{m' \times n}$, any $b \in \mathbb{R}^m$, any $b' \in \mathbb{R}^{m'}$ and any $J \subseteq n$, the linear optimization problem or linear program (LP) in **general form** is

$$\min \left\{ \langle c, x \rangle \;\middle|\; A\,x = b \;\wedge\; A'\,x \geq b' \;\wedge\; (\forall j \in J \colon x_j \geq 0) \;\wedge\; x \in \mathbb{R}^n \right\} \;. \tag{42}$$

Lemma 3. Consider

- any $m, n \in \mathbb{N}_0$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$ and $A \in \mathbb{R}^{m \times n}$ such that $\text{rank}(A) = m$
- any BFS x with basis K to the standard system defined by A and b
- any $k^* \in n$ and the $y^{k^*} \in \mathbb{R}^n$ with $y^{k^*}|_K = A|_{\cdot K}^{-1} A_{\cdot k^*}$ and $y^{k^*}|_{n \setminus K} = 0$
- x' the result of pivoting x in K such that k^* enters the basis.

Then:

$$\langle c, x' \rangle - \langle c, x \rangle = \left(c_{k^*} - \sum_{k \in K} c_k y_k^{k^*} \right) x'_{k^*} \tag{43}$$

Definition 8. Under the assumptions of Lemma 3, the **relative cost** of k^* in the basis K is the number

$$c_{k^*} - \sum_{k \in K} c_k y_k^{k^*} \quad . \tag{44}$$

Proof. If $k^* \in K$, the statement follows from Remark 4c.

Otherwise, $x_{k^*} = 0$. Thus:

$$\langle c, x' \rangle - \langle c, x \rangle = \langle c, x' - x \rangle \tag{45}$$

$$= \sum_{k \in K \cup \{k^*\}} c_k (x'_k - x_k) \tag{46}$$

$$= c_{k^*}(x'_{k^*} - 0) + \sum_{k \in K} c_k (x'_k - x_k) \tag{47}$$

$$\overset{(18)}{=} c_{k^*} x'_{k^*} + \sum_{k \in K} c_k (-x'_{k^*} y_k^{k^*}) \tag{48}$$

$$= \left(c_{k^*} - \sum_{k \in K} c_k y_k^{k^*} \right) x'_{k^*} \ . \tag{49}$$

□

Theorem 3. Let

- $m, n \in \mathbb{N}_0$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$ and $A \in \mathbb{R}^{m \times n}$ such that $\text{rank}(A) = m$
- x a BFS with basis K to the standard system defined by A and b
- for any $k \in n$, c'_k the relative cost of k in the basis K, and $c'_k \geq 0$.

Then, x is a solution to the LP in standard form defined by A and b.

Proof. Let x' be any feasible solution to the standard system defined by A and b, not necessarily basic. Now:

$$
\begin{aligned}
\langle c, x' \rangle &= \sum_{k \in n} c_k x'_k \\
&\geq \sum_{k \in n} \left(\sum_{k' \in K} c_{k'} y^k_{k'} \right) x'_k \\
&= \sum_{k' \in K} c_{k'} \sum_{k \in n} y^k_{k'} x'_k = \sum_{k' \in K} c_{k'} \sum_{k \in n} (A|_{\cdot K}^{-1} A_{\cdot k})_{k'} x'_k \\
&= \sum_{k' \in K} c_{k'} \sum_{k \in n} \left(\sum_{j \in m} (A|_{\cdot K}^{-1})_{k'j} A_{jk} \right) x'_k \\
&= \sum_{k' \in K} c_{k'} \sum_{j \in m} (A|_{\cdot K}^{-1})_{k'j} \sum_{k \in n} A_{jk} x'_k = \sum_{k' \in K} c_{k'} \sum_{j \in m} (A|_{\cdot K}^{-1})_{k'j} (Ax')_j \\
&= \sum_{k' \in K} c_{k'} \sum_{j \in m} (A|_{\cdot K}^{-1})_{k'j} b_j = \sum_{k' \in K} c_{k'} x_{k'} \\
&= \langle c, x \rangle
\end{aligned}
$$

□

Input: a BFS x^0 with basis K

```
t := 0
while true
    for any k ∈ n, let c'_k the relative cost of k in the basis K
    if c' ≥ 0
        return x^t (solution)
    else
        choose k* such that c'_{k*} < 0
        if K_{k*} = ∅
            return unbounded
        else
            choose k† and define λ' according to (16)
            let x^{t+1} the pivoting of x^t s.t. k* enters and k† leaves K
            t := t + 1
```

Simplex Algorithm

Input: a BFS x^0 with basis K

$t := 0$
while true
    for any $k \in n$, let c'_k the relative cost of k in the basis K
    if $c' \geq 0$
        return x^t (solution)
    else
        choose k^* such that $c'_{k^*} < 0$
        if $K_{k^*} = \varnothing$
            return unbounded
        else
            choose $k^\dagger$ and define λ' according to (16)
            let x^{t+1} the pivoting of x^t s.t. k^* enters and $k^\dagger$ leaves K
            $t := t + 1$

Simplex Algorithm

Example 1. Pivoting rules:

– Choose k^* so as to minimize the relative cost

$$c_{k^*} - \sum_{k \in K} c_k y_k^{k^*} \quad . \tag{50}$$

Not guaranteeing locally optimal reduction in cost; not preventing cycling.

– Choose k^* so as to minimize

$$\left(c_{k^*} - \sum_{k \in K} c_k y_k^{k^*} \right) x'_{k^*} \quad . \tag{51}$$

Guaranteeing locally optimal reduction in cost; not preventing cycling.

– Bland's Rule: Define a total order $<$ on K. Among those k^* with a negative reduced cost, choose the one that is smallest wrt. $<$. If there is a tie in (16), choose $k^\dagger$ to be the smallest wrt. $<$.

Not guaranteeing locally optimal reduction in cost; preventing cycling.