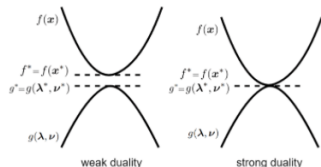
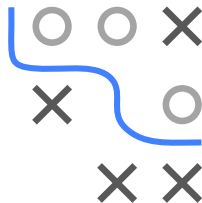


Optimization in Machine Learning

Constrained Optimization

Duality in optimization

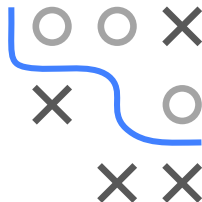


Learning goals

- Awareness of the concept of duality in optimization
- LP duality
- Weak and strong duality in LP

DUALITY: OVERVIEW

- Duality theory plays a fundamental role in (constrained) optimization
- The concept of “duality” emerged in the context of LPs and dates back to the 1940s (works of Tucker and Wolfe)
- There are several different types of duality: LP duality, Lagrangian duality, Wolfe duality, Fenchel duality (which can lead to confusion)
- Key take-home message: The concepts of duality give you recipes to find **lower bounds** on your original “primal” constrained optimization problem
- Under certain conditions, these lower bounds are actually identical to the optimal solution
- Duality is also practical – it has been used to find **better algorithms** for solving constrained optimization problems



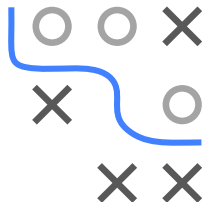
LP DUALITY: INTRODUCTORY EXAMPLE

Example: A bakery sells brownies for 50 ct and mini cheesecakes for 80 ct each

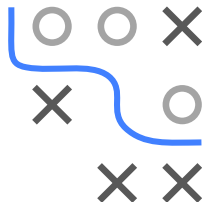
- The two products contain the following ingredients

| | Chocolate | Sugar | Cream cheese |
|------------|-----------|-------|--------------|
| Brownie | 3 | 2 | 2 |
| Cheesecake | 0 | 4 | 5 |

- A student wants to minimize his expenses, but at the same time eat at least 6 units of chocolate, 10 units of sugar and 8 units of cream cheese



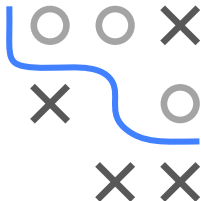
LP DUALITY: INTRODUCTORY EXAMPLE



- He is therefore confronted with the following optimization problem:

$$\begin{aligned} \min_{\mathbf{x} \in \mathbb{R}^2} \quad & 50x_1 + 80x_2 \\ \text{s.t.} \quad & 3x_1 \geq 6, \quad 2x_1 + 4x_2 \geq 10, \quad 2x_1 + 5x_2 \geq 8, \quad \mathbf{x} \geq 0 \end{aligned}$$

LP DUALITY: INTRODUCTORY EXAMPLE



The solution of the Simplex algorithm:

```
res = solveLP(cvec = c, bvec = b, Amat = A)
summary(res)
##
## Results of Linear Programming / Linear Optimization
##
## Objective function (Minimum): 220
##
## Solution
## opt
## 1 2.0
## 2 1.5
```

LP DUALITY: INTRODUCTORY EXAMPLE

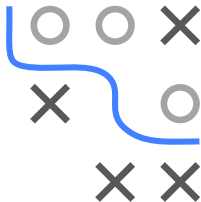
- The baker informs the supplier that he needs at least 6 units of chocolate, 10 units of sugar and 8 units of cream cheese to meet the student's requirements
- The supplier asks himself how he must set the prices for chocolate, sugar and cream cheese $(\alpha_1, \alpha_2, \alpha_3)$ such that he can
 - maximize his revenue

$$\max_{\alpha \in \mathbb{R}^3} 6\alpha_1 + 10\alpha_2 + 8\alpha_3$$

- and at the same time ensure that the baker buys from him (purchase cost \leq selling price)

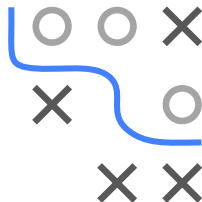
$$3\alpha_1 + 2\alpha_2 + 2\alpha_3 \leq 50 \quad (\text{Brownie})$$

$$4\alpha_2 + 5\alpha_3 \leq 80 \quad (\text{Cheesecake})$$



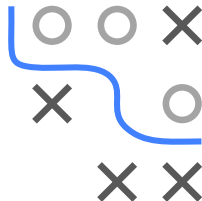
LP DUALITY: INTRODUCTORY EXAMPLE

- The presented example is known as a **dual problem**
- The variables α_i are called **dual variables**
- In an economic context, dual variables can often be interpreted as **shadow prices** for certain goods
- If we solve the dual problem, we see that the dual problem has the same objective function value as the primal problem
- This is later referred to as **strong duality**



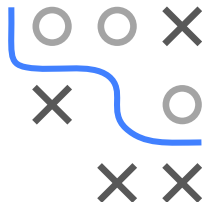
LP DUALITY: INTRODUCTORY EXAMPLE

```
res = solveLP(cvec = c, bvec = b, Amat = A, maximum = T)
summary(res)
##
## Results of Linear Programming / Linear Optimization
##
## Objective function (Maximum): 220
##
## Solution
## opt
## 1 3.333333
## 2 20.000000
## 3 0.000000
```



MATHEMATICAL INTUITION

- The example explained duality from an economic point of view
- But what is the mathematical intuition behind duality?



Idea: In minimization problems one is often interested in **lower bounds** of the objective function

- How could we derive a lower bound for the problem above?
- If we “skillfully” multiply the three inequalities by factors and add factors (similar to a linear system), we can find a lower bound

MATHEMATICAL INTUITION

- Consider the primal problem with multipliers:

$$\min_{\mathbf{x} \in \mathbb{R}^2} 50x_1 + 80x_2$$

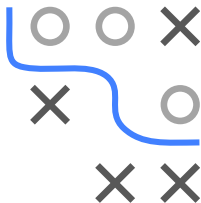
$$\text{s.t. } 3x_1 \geq 6 \mid \cdot 5, \quad 2x_1 + 4x_2 \geq 10 \mid \cdot 5, \quad 2x_1 + 5x_2 \geq 8 \mid \cdot 12, \quad \mathbf{x} \geq 0$$

- If we add up the constraints we obtain

$$5 \cdot (3x_1) + 5 \cdot (2x_1 + 4x_2) + 12 \cdot (2x_1 + 5x_2) = 49x_1 + 80x_2 \geq 30 + 50 + 96 = 176$$

- Since $x_1 \geq 0$ we found a lower bound because

$$50x_1 + 80x_2 \geq 49x_1 + 80x_2 \geq 176$$



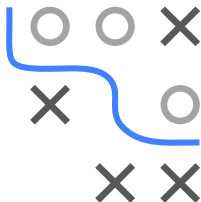
MATHEMATICAL INTUITION

- Is our derived lower bound the best possible?
- We replace the multipliers 5, 5, 12 by $\alpha_1, \alpha_2, \alpha_3$ and compute:

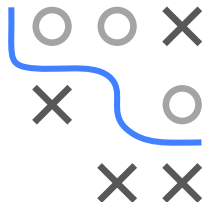
$$\begin{aligned} 50x_1 + 80x_2 &\geq \alpha_1(3x_1) + \alpha_2(2x_1 + 4x_2) + \alpha_3(2x_1 + 5x_2) \\ &= (3\alpha_1 + 2\alpha_2 + 2\alpha_3)x_1 + (4\alpha_2 + 5\alpha_3)x_2 \\ &\geq 6\alpha_1 + 10\alpha_2 + 8\alpha_3 \end{aligned}$$

- **But:** We have to demand that

$$3\alpha_1 + 2\alpha_2 + 2\alpha_3 \leq 50 \quad \text{and} \quad 4\alpha_2 + 5\alpha_3 \leq 80$$



MATHEMATICAL INTUITION

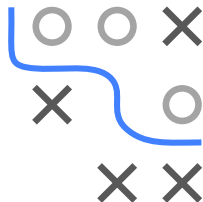


- We are interested in a **largest possible** lower bound
- This yields the **dual problem**:

$$\max_{\alpha \in \mathbb{R}^3} 6\alpha_1 + 10\alpha_2 + 8\alpha_3$$

$$\text{s.t.} \quad 3\alpha_1 + 2\alpha_2 + 2\alpha_3 \leq 50, \quad 4\alpha_2 + 5\alpha_3 \leq 80, \quad \alpha \geq 0$$

DUALITY



Dual problem:

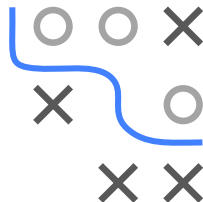
$$\max_{\alpha \in \mathbb{R}^m} g(\alpha) := \alpha^T \mathbf{b} \quad \text{s.t. } \alpha^T \mathbf{A} \leq \mathbf{c}^T, \quad \alpha \geq 0$$

Primal problem:

$$\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}) := \mathbf{c}^T \mathbf{x} \quad \text{s.t. } \mathbf{A}\mathbf{x} \geq \mathbf{b}, \quad \mathbf{x} \geq 0$$

DUALITY

Connection of primal and dual problem:



| | Primal (minimize) | Dual (maximize) | |
|-----------|---------------------------------------|---------------------------------------|-----------|
| condition | \leq \geq $=$ | ≤ 0 ≥ 0 unconstrained | variable |
| variable | ≥ 0 ≤ 0 unconstrained | \leq \geq $=$ | condition |

WEAK DUALITY THEOREM

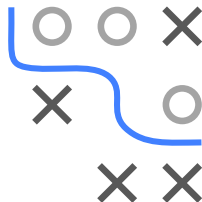
- In general, the **weak duality theorem** applies to all feasible \mathbf{x}, α

$$g(\alpha) = \alpha^T \mathbf{b} \leq \mathbf{c}^T \mathbf{x} = f(\mathbf{x})$$

- The value of the dual function is therefore **always** a lower bound for the objective function value of the primal problem

Proof:

$$\alpha^T \mathbf{b} \stackrel{\mathbf{Ax} \geq \mathbf{b}}{\leq} \alpha^T \mathbf{Ax} \stackrel{\alpha^T \mathbf{A} \leq \mathbf{c}^T}{\leq} \mathbf{c}^T \mathbf{x}$$

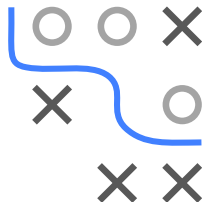


STRONG DUALITY THEOREM

- The **strong duality theorem** states that if one of the two problems has a constrained solution, then the other also has a constrained solution
- The objective function values are the same in this case:

$$g(\alpha^*) = (\alpha^*)^T \mathbf{b} = \mathbf{c}^T \mathbf{x}^* = f(\mathbf{x}^*)$$

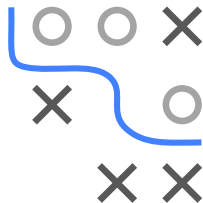
- In this case, the dual problem can be solved instead of the primal problem, which can lead to enormous run time advantages, especially with many constraints and few variables
- The **dual simplex algorithm**, which has emerged as a standard procedure for linear programming, is based on this idea



ALTERNATIVE LP FORMULATION

- Many slightly different (but ultimately equivalent) formulations of primal and dual LPs exist in the literature
- One common alternative with inequality and equality constraints is often formulated as follows
- Let $\mathbf{c} \in \mathbb{R}^d$, $\mathbf{b} \in \mathbb{R}^l$, $\mathbf{A} \in \mathbb{R}^{l \times d}$, $\mathbf{h} \in \mathbb{R}^k$, and $\mathbf{G} \in \mathbb{R}^{k \times d}$
- Then the primal LP is defined as

$$\min_{\mathbf{x} \in \mathbb{R}^d} \mathbf{c}^T \mathbf{x} \quad \text{s.t.} \quad \mathbf{G}\mathbf{x} = \mathbf{h}, \quad \mathbf{A}\mathbf{x} \leq \mathbf{b}$$



ALTERNATIVE LP FORMULATION

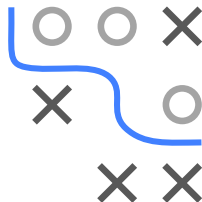
- The corresponding dual LP:

$$\max_{\alpha \in \mathbb{R}^l, \beta \in \mathbb{R}^k} -\mathbf{b}^T \alpha - \mathbf{h}^T \beta \quad \text{s.t.} \quad -\mathbf{A}^T \alpha - \mathbf{G}^T \beta = \mathbf{c}, \quad \alpha \geq 0$$

- The following argument again highlights the interpretation of the dual LP as a lower bound
- Here, for $\alpha \geq 0$ and any β , and \mathbf{x} primal feasible, it holds that

$$\alpha^T (\mathbf{A}\mathbf{x} - \mathbf{b}) + \beta^T (\mathbf{G}\mathbf{x} - \mathbf{h}) \leq 0 \iff (-\mathbf{A}^T \alpha - \mathbf{G}^T \beta)^T \mathbf{x} \geq -\mathbf{b}^T \alpha - \mathbf{h}^T \beta$$

- So if $\mathbf{c} = -\mathbf{A}^T \alpha - \mathbf{G}^T \beta$, we get a lower bound on the primal optimal value



ALTERNATIVE LP FORMULATION

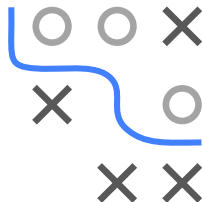
- Another perspective on this formulation will connect LP duality to the more general notion of **Lagrangian duality**
- Again, for $\alpha \geq 0$, any β , and \mathbf{x} primal feasible, it holds that

$$\mathbf{c}^T \mathbf{x} \geq \mathbf{c}^T \mathbf{x} + \alpha^T (\mathbf{A}\mathbf{x} - \mathbf{b}) + \beta^T (\mathbf{G}\mathbf{x} - \mathbf{h}) =: \mathcal{L}(\mathbf{x}, \alpha, \beta)$$

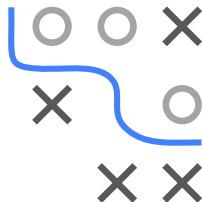
- If \mathcal{S} denotes the primal feasible set, $f(\mathbf{x}^*)$ the primal optimal value, then for $\alpha \geq 0$ and any β , it holds that

$$f(\mathbf{x}^*) \geq \min_{\mathbf{x} \in \mathcal{S}} \mathcal{L}(\mathbf{x}, \alpha, \beta) \geq \min_{\mathbf{x} \in \mathbb{R}^d} \mathcal{L}(\mathbf{x}, \alpha, \beta) =: g(\alpha, \beta)$$

- This shows that $g(\alpha, \beta)$ is a lower bound on $f(\mathbf{x}^*)$ for $\alpha \geq 0$ and any β



ALTERNATIVE LP FORMULATION



- The *Lagrange (dual) function* is defined as

$$g(\alpha, \beta) = \begin{cases} -\mathbf{b}^T \alpha - \mathbf{h}^T \beta & \text{if } \mathbf{c} = -\mathbf{A}^T \alpha - \mathbf{G}^T \beta \\ -\infty & \text{otherwise} \end{cases}$$

- Maximizing $g(\alpha, \beta)$ leads again to the first dual formulation

Note: Lagrangian perspective is completely general \Rightarrow applicable to arbitrary (non-linear) problems

FINAL REMARKS

- We introduced key concepts of duality for Linear Programming as the simplest instance of a constrained optimization problem
- We refer to the excellent course of L. Vandenberghe [▶ Click for source](#) for many more details
- We have skipped algorithmic approaches for solving linear programs: Dantzig's Simplex Algorithm, Interior point methods, and the Ellipsoid method

