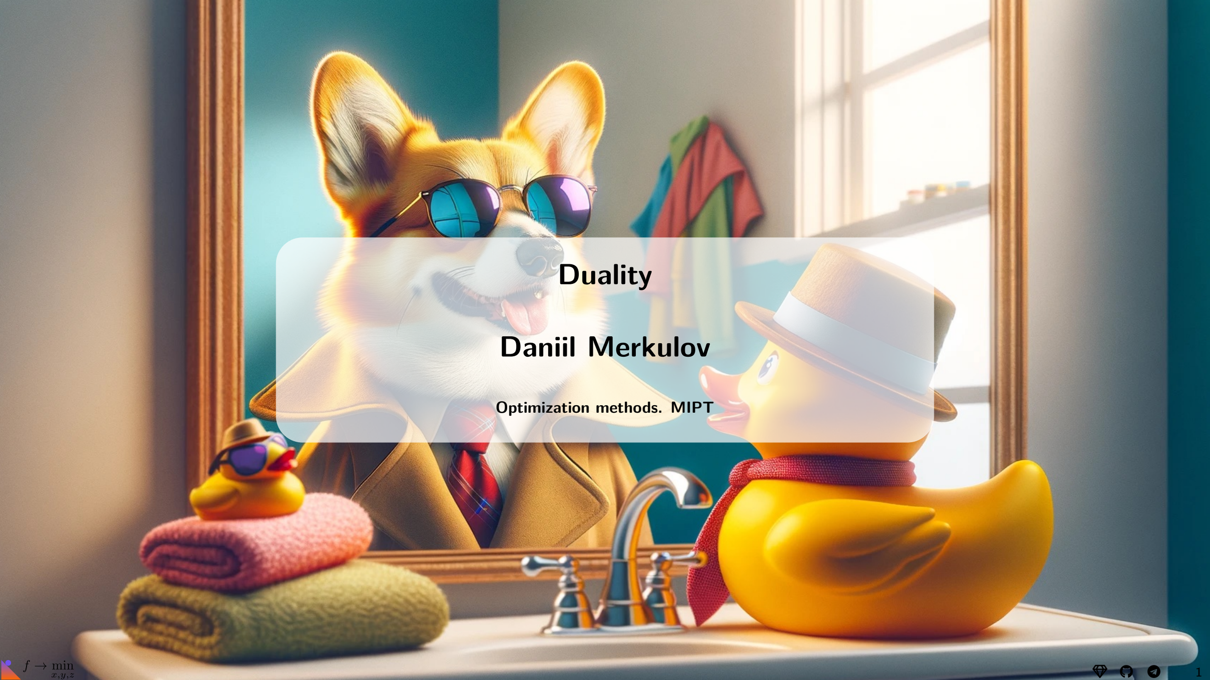

Duality

Daniil Merkulov

Optimization methods. MIPT

Duality

Motivation

Duality lets us associate to any constrained optimization problem a concave maximization problem, whose solutions lower bound the optimal value of the original problem. What is interesting is that there are cases, when one can solve the primal problem by first solving the dual one. Now, consider a general constrained optimization problem:

Motivation

Duality lets us associate to any constrained optimization problem a concave maximization problem, whose solutions lower bound the optimal value of the original problem. What is interesting is that there are cases, when one can solve the primal problem by first solving the dual one. Now, consider a general constrained optimization problem:

$$\text{Primal: } f(x) \to \min_{x \in S} \qquad \text{Dual: } g(y) \to \max_{y \in \Omega}$$

Motivation

Duality lets us associate to any constrained optimization problem a concave maximization problem, whose solutions lower bound the optimal value of the original problem. What is interesting is that there are cases, when one can solve the primal problem by first solving the dual one. Now, consider a general constrained optimization problem:

$$\text{Primal: } f(x) \to \min_{x \in S} \qquad \text{Dual: } g(y) \to \max_{y \in \Omega}$$

We'll build $g(y)$, that preserves the uniform bound:

$$g(y) \leq f(x) \qquad \forall x \in S, \forall y \in \Omega$$

Motivation

Duality lets us associate to any constrained optimization problem a concave maximization problem, whose solutions lower bound the optimal value of the original problem. What is interesting is that there are cases, when one can solve the primal problem by first solving the dual one. Now, consider a general constrained optimization problem:

$$\text{Primal: } f(x) \to \min_{x \in S} \qquad \text{Dual: } g(y) \to \max_{y \in \Omega}$$

We'll build $g(y)$, that preserves the uniform bound:

$$g(y) \leq f(x) \qquad \forall x \in S, \forall y \in \Omega$$

As a consequence:

$$\max_{y \in \Omega} g(y) \leq \min_{x \in S} f(x)$$

Lagrange duality

We'll consider one of many possible ways to construct $g(y)$ in case, when we have a general mathematical programming problem with functional constraints:

Lagrange duality

We'll consider one of many possible ways to construct $g(y)$ in case, when we have a general mathematical programming problem with functional constraints:

$$
\begin{split}
& f_0(x) \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } & f_i(x) \leq 0, \; i = 1, \ldots, m \\
& h_i(x) = 0, \; i = 1, \ldots, p
\end{split}
$$

Lagrange duality

We'll consider one of many possible ways to construct $g(y)$ in case, when we have a general mathematical programming problem with functional constraints:

$$
\begin{split}
& f_0(x) \to \min_{x \in \mathbb{R}^n}\\
\text{s.t. } & f_i(x) \leq 0, \; i = 1, \ldots, m\\
& h_i(x) = 0, \; i = 1, \ldots, p
\end{split}
$$

And the Lagrangian, associated with this problem:

$$
L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x) = f_0(x) + \lambda^\top f(x) + \nu^\top h(x)
$$

Dual function

We assume $\mathcal{D} = \bigcap\limits_{i=0}^{m} \textbf{dom } f_i \cap \bigcap\limits_{i=1}^{p} \textbf{dom } h_i$ is nonempty. We define the Lagrange dual function (or just dual function) $g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ as the minimum value of the Lagrangian over x: for $\lambda \in \mathbb{R}^m, \nu \in \mathbb{R}^p$

Dual function

We assume $\mathcal{D} = \bigcap\limits_{i=0}^{m} \textbf{dom } f_i \cap \bigcap\limits_{i=1}^{p} \textbf{dom } h_i$ is nonempty. We define the Lagrange dual function (or just dual function) $g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ as the minimum value of the Lagrangian over x: for $\lambda \in \mathbb{R}^m, \nu \in \mathbb{R}^p$

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) = \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right)$$

Dual function

We assume $\mathcal{D} = \bigcap\limits_{i=0}^{m} \textbf{dom } f_i \cap \bigcap\limits_{i=1}^{p} \textbf{dom } h_i$ is nonempty. We define the Lagrange dual function (or just dual function) $g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ as the minimum value of the Lagrangian over x: for $\lambda \in \mathbb{R}^m, \nu \in \mathbb{R}^p$

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) = \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right)$$

When the Lagrangian is unbounded below in x, the dual function takes on the value $-\infty$. Since the dual function is the pointwise infimum of a family of affine functions of (λ, ν), it is concave, even when the original problem is not convex.

Dual function as a lower bound

Let us show, that the dual function yields lower bounds on the optimal value p^* of the original problem for any $\lambda \succeq 0, \nu$. Suppose some $\hat{x}$ is a feasible point for the original problem, i.e., $f_i(\hat{x}) \leq 0$ and $h_i(\hat{x}) = 0,\ \lambda \succeq 0$. Then we have:

Dual function as a lower bound

Let us show, that the dual function yields lower bounds on the optimal value p^* of the original problem for any $\lambda \succeq 0, \nu$. Suppose some $\hat{x}$ is a feasible point for the original problem, i.e., $f_i(\hat{x}) \leq 0$ and $h_i(\hat{x}) = 0$, $\lambda \succeq 0$. Then we have:

$$L(\hat{x}, \lambda, \nu) = f_0(\hat{x}) + \underbrace{\lambda^\top f(\hat{x})}_{\leq 0} + \underbrace{\nu^\top h(\hat{x})}_{=0} \leq f_0(\hat{x})$$

Dual function as a lower bound

Let us show, that the dual function yields lower bounds on the optimal value p^* of the original problem for any $\lambda \succeq 0, \nu$. Suppose some $\hat{x}$ is a feasible point for the original problem, i.e., $f_i(\hat{x}) \leq 0$ and $h_i(\hat{x}) = 0,\ \lambda \succeq 0$. Then we have:

$$L(\hat{x}, \lambda, \nu) = f_0(\hat{x}) + \underbrace{\lambda^\top f(\hat{x})}_{\leq 0} + \underbrace{\nu^\top h(\hat{x})}_{=0} \leq f_0(\hat{x})$$

Hence

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \leq L(\hat{x}, \lambda, \nu) \leq f_0(\hat{x})$$

Dual function as a lower bound

Let us show, that the dual function yields lower bounds on the optimal value p^* of the original problem for any $\lambda \succeq 0, \nu$. Suppose some $\hat{x}$ is a feasible point for the original problem, i.e., $f_i(\hat{x}) \leq 0$ and $h_i(\hat{x}) = 0, \; \lambda \succeq 0$. Then we have:

$$
L(\hat{x}, \lambda, \nu) = f_0(\hat{x}) + \underbrace{\lambda^\top f(\hat{x})}_{\leq 0} + \underbrace{\nu^\top h(\hat{x})}_{= 0} \leq f_0(\hat{x})
$$

Hence

$$
g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \leq L(\hat{x}, \lambda, \nu) \leq f_0(\hat{x})
$$

$$
g(\lambda, \nu) \leq p^*
$$

Dual function as a lower bound

Let us show, that the dual function yields lower bounds on the optimal value p^* of the original problem for any $\lambda \succeq 0, \nu$. Suppose some $\hat{x}$ is a feasible point for the original problem, i.e., $f_i(\hat{x}) \leq 0$ and $h_i(\hat{x}) = 0,\ \lambda \succeq 0$. Then we have:

$$L(\hat{x}, \lambda, \nu) = f_0(\hat{x}) + \underbrace{\lambda^\top f(\hat{x})}_{\leq 0} + \underbrace{\nu^\top h(\hat{x})}_{=0} \leq f_0(\hat{x})$$

Hence

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \leq L(\hat{x}, \lambda, \nu) \leq f_0(\hat{x})$$

$$g(\lambda, \nu) \leq p^*$$

A natural question is: what is the *best* lower bound that can be obtained from the Lagrange dual function? This leads to the following optimization problem:

Dual function as a lower bound

Let us show, that the dual function yields lower bounds on the optimal value p^* of the original problem for any $\lambda \succeq 0, \nu$. Suppose some $\hat{x}$ is a feasible point for the original problem, i.e., $f_i(\hat{x}) \leq 0$ and $h_i(\hat{x}) = 0,\ \lambda \succeq 0$. Then we have:

$$L(\hat{x}, \lambda, \nu) = f_0(\hat{x}) + \underbrace{\lambda^\top f(\hat{x})}_{\leq 0} + \underbrace{\nu^\top h(\hat{x})}_{=0} \leq f_0(\hat{x})$$

Hence

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \leq L(\hat{x}, \lambda, \nu) \leq f_0(\hat{x})$$

$$g(\lambda, \nu) \leq p^*$$

A natural question is: what is the *best* lower bound that can be obtained from the Lagrange dual function? This leads to the following optimization problem:

$$g(\lambda, \nu) \to \max_{\lambda \in \mathbb{R}^m,\ \nu \in \mathbb{R}^p}$$

$$\text{s.t. } \lambda \succeq 0$$

Dual function as a lower bound

Let us show, that the dual function yields lower bounds on the optimal value p^* of the original problem for any $\lambda \succeq 0, \nu$. Suppose some $\hat{x}$ is a feasible point for the original problem, i.e., $f_i(\hat{x}) \leq 0$ and $h_i(\hat{x}) = 0,\ \lambda \succeq 0$. Then we have:

$$
L(\hat{x}, \lambda, \nu) = f_0(\hat{x}) + \underbrace{\lambda^\top f(\hat{x})}_{\leq 0} + \underbrace{\nu^\top h(\hat{x})}_{=0} \leq f_0(\hat{x})
$$

Hence

$$
g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \leq L(\hat{x}, \lambda, \nu) \leq f_0(\hat{x})
$$

$$
g(\lambda, \nu) \leq p^*
$$

A natural question is: what is the *best* lower bound that can be obtained from the Lagrange dual function? This leads to the following optimization problem:

$$
g(\lambda, \nu) \to \max_{\lambda \in \mathbb{R}^m, \ \nu \in \mathbb{R}^p}
$$

$$
\text{s.t. } \lambda \succeq 0
$$

The term "dual feasible", to describe a pair (λ, ν) with $\lambda \succeq 0$ and $g(\lambda, \nu) > -\infty$, now makes sense. It means, as the name implies, that (λ, ν) is feasible for the dual problem. We refer to (λ^*, ν^*) as dual optimal or optimal Lagrange multipliers if they are optimal for the above problem.

Summary

	Primal	Dual
Function	$f_0(x)$	$g(\lambda, \nu) = \min\limits_{x \in \mathcal{D}} L(x, \lambda, \nu)$
Variables	$x \in S \subseteq \mathbb{R}^n$	$\lambda \in \mathbb{R}^m_+, \nu \in \mathbb{R}^p$
Constraints	$f_i(x) \le 0,\ i = 1, \ldots, m$ $h_i(x) = 0,\ i = 1, \ldots, p$	$\lambda_i \ge 0, \forall i \in \overline{1, m}$
Problem	$f_0(x) \to \min\limits_{x \in \mathbb{R}^n}$ s.t. $f_i(x) \le 0,\ i = 1, \ldots, m$ $h_i(x) = 0,\ i = 1, \ldots, p$	$g(\lambda, \nu) \to \max\limits_{\lambda \in \mathbb{R}^m, \nu \in \mathbb{R}^p}$ s.t. $\lambda \succeq 0$
Optimal	x^* if feasible, $p^* = f_0(x^*)$	λ^*, ν^* if $\max$ is achieved, $d^* = g(\lambda^*, \nu^*)$

Example. Linear Least Squares

We are addressing a problem within a non-empty budget set, defined as follows:

Example. Linear Least Squares

We are addressing a problem within a non-empty budget set, defined as follows:

$$\begin{aligned} \text{min} \quad & x^T x \\ \text{s.t.} \quad & Ax = b, \end{aligned}$$

with the matrix $A \in \mathbb{R}^{m \times n}$.

Example. Linear Least Squares

We are addressing a problem within a non-empty budget set, defined as follows:

$$
\begin{aligned}
&\text{min} \quad x^T x \\
&\text{s.t.} \quad Ax = b,
\end{aligned}
$$

with the matrix $A \in \mathbb{R}^{m \times n}$.

This problem is devoid of inequality constraints, presenting m linear equality constraints instead. The Lagrangian is expressed as $L(x, \nu) = x^T x + \nu^T (Ax - b)$, spanning the domain $\mathbb{R}^n \times \mathbb{R}^m$. The dual function is denoted by $g(\nu) = \inf_x L(x, \nu)$. Given that $L(x, \nu)$ manifests as a convex quadratic function in terms of x, the minimizing x can be derived from the optimality condition

Example. Linear Least Squares

We are addressing a problem within a non-empty budget set, defined as follows:

$$
\begin{aligned}
&\text{min} \quad x^T x \\
&\text{s.t.} \quad Ax = b,
\end{aligned}
$$

with the matrix $A \in \mathbb{R}^{m \times n}$.

This problem is devoid of inequality constraints, presenting m linear equality constraints instead. The Lagrangian is expressed as $L(x, \nu) = x^T x + \nu^T (Ax - b)$, spanning the domain $\mathbb{R}^n \times \mathbb{R}^m$. The dual function is denoted by $g(\nu) = \inf_x L(x, \nu)$. Given that $L(x, \nu)$ manifests as a convex quadratic function in terms of x, the minimizing x can be derived from the optimality condition

$$
\nabla_x L(x, \nu) = 2x + A^T \nu = 0,
$$

Example. Linear Least Squares

We are addressing a problem within a non-empty budget set, defined as follows:

$$
\begin{aligned}
\text{min} \quad & x^T x \\
\text{s.t.} \quad & Ax = b,
\end{aligned}
$$

with the matrix $A \in \mathbb{R}^{m \times n}$.

This problem is devoid of inequality constraints, presenting m linear equality constraints instead. The Lagrangian is expressed as $L(x, \nu) = x^T x + \nu^T (Ax - b)$, spanning the domain $\mathbb{R}^n \times \mathbb{R}^m$. The dual function is denoted by $g(\nu) = \inf_x L(x, \nu)$. Given that $L(x, \nu)$ manifests as a convex quadratic function in terms of x, the minimizing x can be derived from the optimality condition

$$
\nabla_x L(x, \nu) = 2x + A^T \nu = 0,
$$

leading to $x = -(1/2)A^T \nu$. As a result, the dual function is articulated as

Example. Linear Least Squares

We are addressing a problem within a non-empty budget set, defined as follows:

$$\begin{aligned} \text{min} \quad & x^T x \\ \text{s.t.} \quad & Ax = b, \end{aligned}$$

with the matrix $A \in \mathbb{R}^{m \times n}$.

This problem is devoid of inequality constraints, presenting m linear equality constraints instead. The Lagrangian is expressed as $L(x, \nu) = x^T x + \nu^T (Ax - b)$, spanning the domain $\mathbb{R}^n \times \mathbb{R}^m$. The dual function is denoted by $g(\nu) = \inf_x L(x, \nu)$. Given that $L(x, \nu)$ manifests as a convex quadratic function in terms of x, the minimizing x can be derived from the optimality condition

$$\nabla_x L(x, \nu) = 2x + A^T \nu = 0,$$

leading to $x = -(1/2)A^T \nu$. As a result, the dual function is articulated as

$$g(\nu) = L(-(1/2)A^T \nu, \nu) = -(1/4)\nu^T A A^T \nu - b^T \nu,$$

Example. Linear Least Squares

We are addressing a problem within a non-empty budget set, defined as follows:

$$
\begin{aligned}
\text{min} \quad & x^T x \\
\text{s.t.} \quad & Ax = b,
\end{aligned}
$$

with the matrix $A \in \mathbb{R}^{m \times n}$.

This problem is devoid of inequality constraints, presenting m linear equality constraints instead. The Lagrangian is expressed as $L(x, \nu) = x^T x + \nu^T (Ax - b)$, spanning the domain $\mathbb{R}^n \times \mathbb{R}^m$. The dual function is denoted by $g(\nu) = \inf_x L(x, \nu)$. Given that $L(x, \nu)$ manifests as a convex quadratic function in terms of x, the minimizing x can be derived from the optimality condition

$$
\nabla_x L(x, \nu) = 2x + A^T \nu = 0,
$$

leading to $x = -(1/2)A^T \nu$. As a result, the dual function is articulated as

$$
g(\nu) = L(-(1/2)A^T \nu, \nu) = -(1/4)\nu^T A A^T \nu - b^T \nu,
$$

emerging as a concave quadratic function within the domain $\mathbb{R}^p$. According to the lower bound property, for any $\nu \in \mathbb{R}^p$, the following holds true:

Example. Linear Least Squares

We are addressing a problem within a non-empty budget set, defined as follows:

$$
\begin{aligned}
\text{min} \quad & x^T x \\
\text{s.t.} \quad & Ax = b,
\end{aligned}
$$

with the matrix $A \in \mathbb{R}^{m \times n}$.

This problem is devoid of inequality constraints, presenting m linear equality constraints instead. The Lagrangian is expressed as $L(x, \nu) = x^T x + \nu^T (Ax - b)$, spanning the domain $\mathbb{R}^n \times \mathbb{R}^m$. The dual function is denoted by $g(\nu) = \inf_x L(x, \nu)$. Given that $L(x, \nu)$ manifests as a convex quadratic function in terms of x, the minimizing x can be derived from the optimality condition

$$
\nabla_x L(x, \nu) = 2x + A^T \nu = 0,
$$

leading to $x = -(1/2)A^T \nu$. As a result, the dual function is articulated as

$$
g(\nu) = L(-(1/2)A^T \nu, \nu) = -(1/4)\nu^T A A^T \nu - b^T \nu,
$$

emerging as a concave quadratic function within the domain $\mathbb{R}^p$. According to the lower bound property, for any $\nu \in \mathbb{R}^p$, the following holds true:

$$
-(1/4)\nu^T A A^T \nu - b^T \nu \leq \inf\{x^T x \mid Ax = b\}.
$$

Which is a simple non-trivial lower bound without any problem solving.

Example. Two-way partitioning problem

We are examining a (nonconvex) problem:

$$
\begin{aligned}
\text{minimize} \quad & x^T W x \\
\text{subject to} \quad & x_i^2 = 1, \quad i = 1, \ldots, n,
\end{aligned}
$$

Example. Two-way partitioning problem

We are examining a (nonconvex) problem:

$$
\begin{array}{ll}
\text{minimize} & x^T W x \\
\text{subject to} & x_i^2 = 1, \quad i = 1, \dots, n,
\end{array}
$$

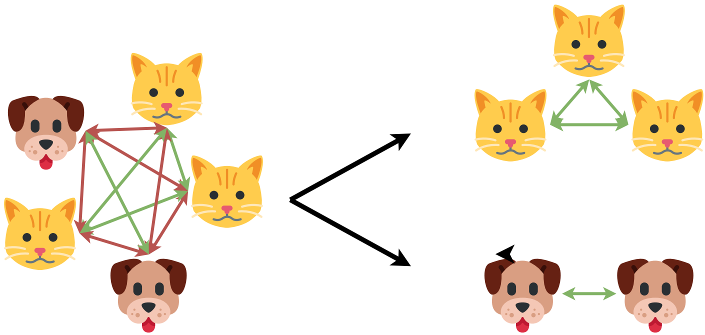

Figure 1: Illustration of two-way partitioning problem

Example. Two-way partitioning problem

We are examining a (nonconvex) problem:

$$
\begin{aligned}
\text{minimize} \quad & x^T W x \\
\text{subject to} \quad & x_i^2 = 1, \quad i = 1, \ldots, n,
\end{aligned}
$$

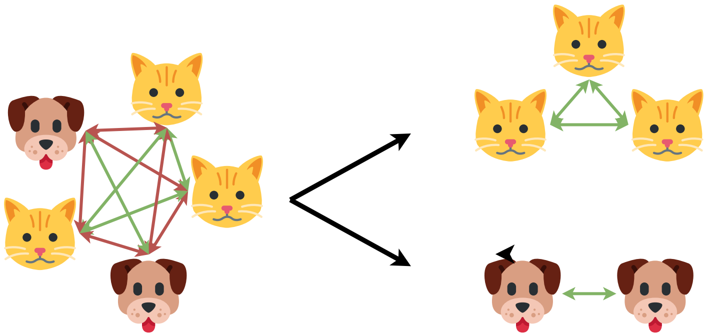

Figure 1: Illustration of two-way partitioning problem

This problem can be construed as a two-way partitioning problem over a set of n elements, denoted as $\{1, \ldots, n\}$: A viable x corresponds to the partition

$$
\{1, \ldots, n\} = \{i | x_i = -1\} \cup \{i | x_i = 1\}.
$$

Example. Two-way partitioning problem

We are examining a (nonconvex) problem:

$$
\begin{aligned}
&\text{minimize} \quad x^T W x \\
&\text{subject to} \quad x_i^2 = 1, \quad i = 1, \ldots, n,
\end{aligned}
$$

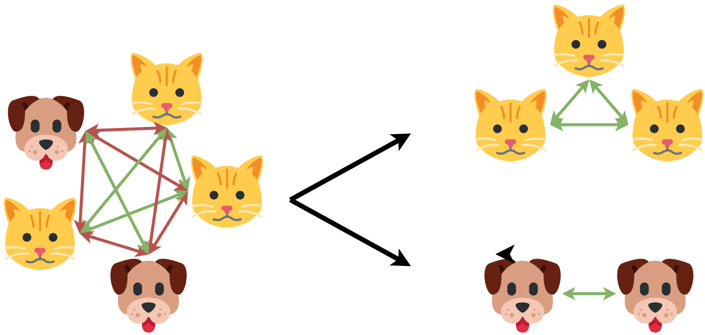

Figure 1: Illustration of two-way partitioning problem

This problem can be construed as a two-way partitioning problem over a set of n elements, denoted as $\{1, \ldots, n\}$: A viable x corresponds to the partition

$$
\{1, \ldots, n\} = \{i | x_i = -1\} \cup \{i | x_i = 1\}.
$$

The coefficient W_{ij} in the matrix represents the expense associated with placing elements i and j in the same partition, while $-W_{ij}$ signifies the cost of segregating them. The objective encapsulates the aggregate cost across all pairs of elements, and the challenge posed by problem is to find the partition that minimizes the total cost.

Example. Two-way partitioning problem

We now derive the dual function for this problem. The Lagrangian is expressed as

$$L(x, \nu) = x^T W x + \sum_{i=1}^{n} \nu_i (x_i^2 - 1) = x^T (W + \mathrm{diag}(\nu)) x - \mathbf{1}^T \nu.$$

Example. Two-way partitioning problem

We now derive the dual function for this problem. The Lagrangian is expressed as

$$L(x, \nu) = x^T W x + \sum_{i=1}^{n} \nu_i (x_i^2 - 1) = x^T (W + \text{diag}(\nu)) x - \mathbf{1}^T \nu.$$

By minimizing over x, we procure the Lagrange dual function:

$$g(\nu) = \inf_x x^T (W + \text{diag}(\nu)) x - \mathbf{1}^T \nu = \begin{cases} -\mathbf{1}^T \nu & \text{if } W + \text{diag}(\nu) \succeq 0 \\ -\infty & \text{otherwise,} \end{cases}$$

Example. Two-way partitioning problem

We now derive the dual function for this problem. The Lagrangian is expressed as

$$L(x, \nu) = x^T W x + \sum_{i=1}^n \nu_i (x_i^2 - 1) = x^T (W + \text{diag}(\nu)) x - \mathbf{1}^T \nu.$$

By minimizing over x, we procure the Lagrange dual function:

$$g(\nu) = \inf_x x^T (W + \text{diag}(\nu)) x - \mathbf{1}^T \nu = \begin{cases} -\mathbf{1}^T \nu & \text{if } W + \text{diag}(\nu) \succeq 0 \\ -\infty & \text{otherwise,} \end{cases}$$

exploiting the realization that the infimum of a quadratic form is either zero (when the form is positive semidefinite) or $-\infty$ (when it's not).

Example. Two-way partitioning problem

We now derive the dual function for this problem. The Lagrangian is expressed as

$$L(x,\nu) = x^T W x + \sum_{i=1}^n \nu_i (x_i^2 - 1) = x^T (W + \textsf{diag}(\nu))x - \mathbf{1}^T \nu.$$

By minimizing over x, we procure the Lagrange dual function:

$$g(\nu) = \inf_x x^T (W + \textsf{diag}(\nu))x - \mathbf{1}^T \nu = \begin{cases} -\mathbf{1}^T \nu & \text{if } W + \textsf{diag}(\nu) \succeq 0 \\ -\infty & \text{otherwise,} \end{cases}$$

exploiting the realization that the infimum of a quadratic form is either zero (when the form is positive semidefinite) or $-\infty$ (when it's not).

This dual function furnishes lower bounds on the optimal value of the problem. For instance, we can adopt the particular value of the dual variable

$$\nu = -\lambda_{\textsf{min}}(W)\mathbf{1}$$

Example. Two-way partitioning problem

We now derive the dual function for this problem. The Lagrangian is expressed as

$$L(x, \nu) = x^T W x + \sum_{i=1}^{n} \nu_i (x_i^2 - 1) = x^T (W + \text{diag}(\nu)) x - \mathbf{1}^T \nu.$$

By minimizing over x, we procure the Lagrange dual function:

$$g(\nu) = \inf_x x^T (W + \text{diag}(\nu)) x - \mathbf{1}^T \nu = \begin{cases} -\mathbf{1}^T \nu & \text{if } W + \text{diag}(\nu) \succeq 0 \\ -\infty & \text{otherwise,} \end{cases}$$

exploiting the realization that the infimum of a quadratic form is either zero (when the form is positive semidefinite) or $-\infty$ (when it's not).

This dual function furnishes lower bounds on the optimal value of the problem. For instance, we can adopt the particular value of the dual variable

$$\nu = -\lambda_{\text{min}}(W) \mathbf{1}$$

which is dual feasible, since $W + \text{diag}(\nu) = W - \lambda_{\text{min}}(W) I \succeq 0$.

Example. Two-way partitioning problem

We now derive the dual function for this problem. The Lagrangian is expressed as

$$L(x,\nu) = x^T W x + \sum_{i=1}^n \nu_i(x_i^2 - 1) = x^T(W + \text{diag}(\nu))x - \mathbf{1}^T\nu.$$

By minimizing over x, we procure the Lagrange dual function:

$$g(\nu) = \inf_x x^T(W + \text{diag}(\nu))x - \mathbf{1}^T\nu = \begin{cases} -\mathbf{1}^T\nu & \text{if } W + \text{diag}(\nu) \succeq 0 \\ -\infty & \text{otherwise,} \end{cases}$$

exploiting the realization that the infimum of a quadratic form is either zero (when the form is positive semidefinite) or $-\infty$ (when it's not).

This dual function furnishes lower bounds on the optimal value of the problem. For instance, we can adopt the particular value of the dual variable

$$\nu = -\lambda_{\text{min}}(W)\mathbf{1}$$

which is dual feasible, since $W + \text{diag}(\nu) = W - \lambda_{\text{min}}(W)I \succeq 0$.

This renders a simple bound on the optimal value p^*: $p^* \geq -\mathbf{1}^T\nu = n\lambda_{\text{min}}(W)$.

Example. Two-way partitioning problem

We now derive the dual function for this problem. The Lagrangian is expressed as

$$L(x,\nu) = x^T W x + \sum_{i=1}^{n} \nu_i (x_i^2 - 1) = x^T (W + \mathbf{diag}(\nu))x - \mathbf{1}^T \nu.$$

By minimizing over x, we procure the Lagrange dual function:

$$g(\nu) = \inf_x x^T (W + \mathbf{diag}(\nu))x - \mathbf{1}^T \nu = \begin{cases} -\mathbf{1}^T \nu & \text{if } W + \mathbf{diag}(\nu) \succeq 0 \\ -\infty & \text{otherwise,} \end{cases}$$

exploiting the realization that the infimum of a quadratic form is either zero (when the form is positive semidefinite) or $-\infty$ (when it's not).

This dual function furnishes lower bounds on the optimal value of the problem. For instance, we can adopt the particular value of the dual variable

$$\nu = -\lambda_{\min}(W)\mathbf{1}$$

which is dual feasible, since $W + \mathbf{diag}(\nu) = W - \lambda_{\min}(W)I \succeq 0$.

This renders a simple bound on the optimal value p^*: $p^* \geq -\mathbf{1}^T \nu = n\lambda_{\min}(W)$.

The code for the problem is available here Open in Colab

Strong duality

Strong duality

It is common to name this relation between optimals of primal and dual problems as **weak duality**. For problem, we have:

$$
p^* \geq d^*
$$

Strong duality

It is common to name this relation between optimals of primal and dual problems as **weak duality**. For problem, we have:

$$p^* \geq d^*$$

While the difference between them is often called **duality gap:**

$$p^* - d^* \geq 0$$

Strong duality

It is common to name this relation between optimals of primal and dual problems as **weak duality**. For problem, we have:

$$
p^* \geq d^*
$$

While the difference between them is often called **duality gap:**

$$
p^* - d^* \geq 0
$$

Note, that we always have weak duality, if we've formulated primal and dual problem. It means, that if we have managed to solve the dual problem (which is always concave, no matter whether the initial problem was or not), then we have some lower bound. Surprisingly, there are some notable cases, when these solutions are equal.

Strong duality

It is common to name this relation between optimals of primal and dual problems as **weak duality**. For problem, we have:

$$
p^* \geq d^*
$$

While the difference between them is often called **duality gap:**

$$
p^* - d^* \geq 0
$$

Note, that we always have weak duality, if we've formulated primal and dual problem. It means, that if we have managed to solve the dual problem (which is always concave, no matter whether the initial problem was or not), then we have some lower bound. Surprisingly, there are some notable cases, when these solutions are equal.

Strong duality happens if duality gap is zero:

$$
p^* = d^*
$$

Strong duality

It is common to name this relation between optimals of primal and dual problems as **weak duality**. For problem, we have:

$$
p^* \geq d^*
$$

While the difference between them is often called **duality gap:**

$$
p^* - d^* \geq 0
$$

Note, that we always have weak duality, if we've formulated primal and dual problem. It means, that if we have managed to solve the dual problem (which is always concave, no matter whether the initial problem was or not), then we have some lower bound. Surprisingly, there are some notable cases, when these solutions are equal.

Strong duality happens if duality gap is zero:

$$
p^* = d^*
$$

Notice: both p^* and d^* may be ∞.

- Several sufficient conditions known!

Strong duality

It is common to name this relation between optimals of primal and dual problems as **weak duality**. For problem, we have:

$$
p^* \geq d^*
$$

While the difference between them is often called **duality gap:**

$$
p^* - d^* \geq 0
$$

Note, that we always have weak duality, if we've formulated primal and dual problem. It means, that if we have managed to solve the dual problem (which is always concave, no matter whether the initial problem was or not), then we have some lower bound. Surprisingly, there are some notable cases, when these solutions are equal.

Strong duality happens if duality gap is zero:

$$
p^* = d^*
$$

Notice: both p^* and d^* may be ∞.

- Several sufficient conditions known!
- "Easy" necessary and sufficient conditions: unknown.

Strong duality in linear least squares

Exercise

In the Least-squares solution of linear equations example above calculate the primal optimum p^* and the dual optimum d^* and check whether this problem has strong duality or not.

Useful features of duality

- **Construction of lower bound on solution of the primal problem.**

  It could be very complicated to solve the initial problem. But if we have the dual problem, we can take an arbitrary $y \in \Omega$ and substitute it in $g(y)$ - we'll immediately obtain some lower bound.

Useful features of duality

- **Construction of lower bound on solution of the primal problem.**

  It could be very complicated to solve the initial problem. But if we have the dual problem, we can take an arbitrary $y \in \Omega$ and substitute it in $g(y)$ - we'll immediately obtain some lower bound.

- **Checking for the problem's solvability and attainability of the solution.**

  From the inequality $\max\limits_{y \in \Omega} g(y) \leq \min\limits_{x \in S} f_0(x)$ follows: if $\min\limits_{x \in S} f_0(x) = -\infty$, then $\Omega = \varnothing$ and vice versa.

Useful features of duality

- **Construction of lower bound on solution of the primal problem.**

  It could be very complicated to solve the initial problem. But if we have the dual problem, we can take an arbitrary $y \in \Omega$ and substitute it in $g(y)$ - we'll immediately obtain some lower bound.

- **Checking for the problem's solvability and attainability of the solution.**

  From the inequality $\max\limits_{y \in \Omega} g(y) \leq \min\limits_{x \in S} f_0(x)$ follows: if $\min\limits_{x \in S} f_0(x) = -\infty$, then $\Omega = \varnothing$ and vice versa.

- **Sometimes it is easier to solve a dual problem than a primal one.**

  In this case, if the strong duality holds: $g(y^*) = f_0(x^*)$ we lose nothing.

Useful features of duality

- **Construction of lower bound on solution of the primal problem.**

  It could be very complicated to solve the initial problem. But if we have the dual problem, we can take an arbitrary $y \in \Omega$ and substitute it in $g(y)$ - we'll immediately obtain some lower bound.

- **Checking for the problem's solvability and attainability of the solution.**

  From the inequality $\max\limits_{y \in \Omega} g(y) \leq \min\limits_{x \in S} f_0(x)$ follows: if $\min\limits_{x \in S} f_0(x) = -\infty$, then $\Omega = \varnothing$ and vice versa.

- **Sometimes it is easier to solve a dual problem than a primal one.**

  In this case, if the strong duality holds: $g(y^*) = f_0(x^*)$ we lose nothing.

- **Obtaining a lower bound on the function's residual.**

  $f_0(x) - f_0^* \leq f_0(x) - g(y)$ for an arbitrary $y \in \Omega$ (suboptimality certificate). Moreover, $p^* \in [g(y), f_0(x)], d^* \in [g(y), f_0(x)]$

Useful features of duality

- **Construction of lower bound on solution of the primal problem.**

  It could be very complicated to solve the initial problem. But if we have the dual problem, we can take an arbitrary $y \in \Omega$ and substitute it in $g(y)$ - we'll immediately obtain some lower bound.

- **Checking for the problem's solvability and attainability of the solution.**

  From the inequality $\max\limits_{y \in \Omega} g(y) \leq \min\limits_{x \in S} f_0(x)$ follows: if $\min\limits_{x \in S} f_0(x) = -\infty$, then $\Omega = \varnothing$ and vice versa.

- **Sometimes it is easier to solve a dual problem than a primal one.**

  In this case, if the strong duality holds: $g(y^*) = f_0(x^*)$ we lose nothing.

- **Obtaining a lower bound on the function's residual.**

  $f_0(x) - f_0^* \leq f_0(x) - g(y)$ for an arbitrary $y \in \Omega$ (suboptimality certificate). Moreover, $p^* \in [g(y), f_0(x)], d^* \in [g(y), f_0(x)]$

- **Dual function is always concave**

  As a pointwise minimum of affine functions.

Slater's condition

Theorem

If for a convex optimization problem (i.e., assuming minimization, f_0, f_i are convex and h_i are affine), there exists a point x such that $h(x) = 0$ and $f_i(x) < 0$ (existance of a strictly feasible point), then we have a zero duality gap and KKT conditions become necessary and sufficient.

An example of convex problem, when Slater's condition does not hold

Example

$$\min\{f_0(x) = x \mid f_1(x) = \frac{x^2}{2} \leq 0\},$$

An example of convex problem, when Slater's condition does not hold

Example

$$
\min\{f_0(x) = x \mid f_1(x) = \frac{x^2}{2} \leq 0\},
$$

The only point in the budget set is: $x^* = 0$. However, it is impossible to find a non-negative $\lambda^* \geq 0$, such that

$$
\nabla f_0(0) + \lambda^* \nabla f_1(0) = 1 + \lambda^* x = 0.
$$

A nonconvex quadratic problem with strong duality

On rare occasions strong duality obtains for a nonconvex problem. As an important example, we consider the problem of minimizing a nonconvex quadratic function over the unit ball

A nonconvex quadratic problem with strong duality

On rare occasions strong duality obtains for a nonconvex problem. As an important example, we consider the problem of minimizing a nonconvex quadratic function over the unit ball

$$
\begin{aligned}
& x^\top A x + 2b^\top x \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } & x^\top x \leq 1
\end{aligned}
$$

A nonconvex quadratic problem with strong duality

On rare occasions strong duality obtains for a nonconvex problem. As an important example, we consider the problem of minimizing a nonconvex quadratic function over the unit ball

$$
\begin{aligned}
&x^\top A x + 2b^\top x \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } &x^\top x \leq 1
\end{aligned}
$$

where $A \in \mathbb{S}^n, A \nsucceq 0$ and $b \in \mathbb{R}^n$. Since $A \nsucceq 0$, this is not a convex problem. This problem is sometimes called the trust region problem, and arises in minimizing a second-order approximation of a function over the unit ball, which is the region in which the approximation is assumed to be approximately valid.

A nonconvex quadratic problem with strong duality

On rare occasions strong duality obtains for a nonconvex problem. As an important example, we consider the problem of minimizing a nonconvex quadratic function over the unit ball

Solution

$$
\begin{aligned}
& x^\top A x + 2b^\top x \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } & x^\top x \leq 1
\end{aligned}
$$

where $A \in \mathbb{S}^n, A \nsucceq 0$ and $b \in \mathbb{R}^n$. Since $A \nsucceq 0$, this is not a convex problem. This problem is sometimes called the trust region problem, and arises in minimizing a second-order approximation of a function over the unit ball, which is the region in which the approximation is assumed to be approximately valid.

A nonconvex quadratic problem with strong duality

On rare occasions strong duality obtains for a nonconvex problem. As an important example, we consider the problem of minimizing a nonconvex quadratic function over the unit ball

$$
\begin{aligned}
& x^\top A x + 2b^\top x \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } & x^\top x \leq 1
\end{aligned}
$$

where $A \in \mathbb{S}^n, A \nsucceq 0$ and $b \in \mathbb{R}^n$. Since $A \nsucceq 0$, this is not a convex problem. This problem is sometimes called the trust region problem, and arises in minimizing a second-order approximation of a function over the unit ball, which is the region in which the approximation is assumed to be approximately valid.

Solution

Lagrangian and dual function

$$
L(x, \lambda) = x^\top A x + 2b^\top x + \lambda(x^\top x - 1) = x^\top (A + \lambda I)x + 2b^\top x - \lambda
$$

A nonconvex quadratic problem with strong duality

On rare occasions strong duality obtains for a nonconvex problem. As an important example, we consider the problem of minimizing a nonconvex quadratic function over the unit ball

$$
\begin{split}
& x^\top A x + 2b^\top x \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } & x^\top x \leq 1
\end{split}
$$

where $A \in \mathbb{S}^n, A \nsucceq 0$ and $b \in \mathbb{R}^n$. Since $A \nsucceq 0$, this is not a convex problem. This problem is sometimes called the trust region problem, and arises in minimizing a second-order approximation of a function over the unit ball, which is the region in which the approximation is assumed to be approximately valid.

Solution

Lagrangian and dual function

$$
L(x, \lambda) = x^\top A x + 2b^\top x + \lambda(x^\top x - 1) = x^\top (A + \lambda I)x + 2b^\top x - \lambda
$$

$$
g(\lambda) = \begin{cases} -b^\top (A + \lambda I)^\dagger b - \lambda & \text{if } A + \lambda I \succeq 0 \\ -\infty, & \text{otherwise} \end{cases}
$$

A nonconvex quadratic problem with strong duality

On rare occasions strong duality obtains for a nonconvex problem. As an important example, we consider the problem of minimizing a nonconvex quadratic function over the unit ball

$$
\begin{split}
& x^\top A x + 2b^\top x \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } & x^\top x \leq 1
\end{split}
$$

where $A \in \mathbb{S}^n, A \nsucceq 0$ and $b \in \mathbb{R}^n$. Since $A \nsucceq 0$, this is not a convex problem. This problem is sometimes called the trust region problem, and arises in minimizing a second-order approximation of a function over the unit ball, which is the region in which the approximation is assumed to be approximately valid.

Solution

Lagrangian and dual function

$$
L(x, \lambda) = x^\top A x + 2b^\top x + \lambda(x^\top x - 1) = x^\top (A + \lambda I) x + 2b^\top x - \lambda
$$

$$
g(\lambda) = \begin{cases} -b^\top (A + \lambda I)^\dagger b - \lambda & \text{if } A + \lambda I \succeq 0 \\ -\infty, & \text{otherwise} \end{cases}
$$

Dual problem:

$$
\begin{split}
& -b^\top (A + \lambda I)^\dagger b - \lambda \to \max_{\lambda \in \mathbb{R}} \\
\text{s.t. } & A + \lambda I \succeq 0
\end{split}
$$

A nonconvex quadratic problem with strong duality

On rare occasions strong duality obtains for a nonconvex problem. As an important example, we consider the problem of minimizing a nonconvex quadratic function over the unit ball

$$
\begin{split}
& x^\top A x + 2b^\top x \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } & x^\top x \leq 1
\end{split}
$$

where $A \in \mathbb{S}^n, A \nsucceq 0$ and $b \in \mathbb{R}^n$. Since $A \nsucceq 0$, this is not a convex problem. This problem is sometimes called the trust region problem, and arises in minimizing a second-order approximation of a function over the unit ball, which is the region in which the approximation is assumed to be approximately valid.

Solution

Lagrangian and dual function

$$
L(x, \lambda) = x^\top A x + 2b^\top x + \lambda(x^\top x - 1) = x^\top (A + \lambda I)x + 2b^\top x - \lambda
$$

$$
g(\lambda) = \begin{cases} -b^\top (A + \lambda I)^\dagger b - \lambda & \text{if } A + \lambda I \succeq 0 \\ -\infty, & \text{otherwise} \end{cases}
$$

Dual problem:

$$
\begin{split}
& -b^\top (A + \lambda I)^\dagger b - \lambda \to \max_{\lambda \in \mathbb{R}} \\
\text{s.t. } & A + \lambda I \succeq 0
\end{split}
$$

$$
\begin{split}
& -\sum_{i=1}^n \frac{(q_i^\top b)^2}{\lambda_i + \lambda} - \lambda \to \max_{\lambda \in \mathbb{R}} \\
\text{s.t. } & \lambda \geq -\lambda_{min}(A)
\end{split}
$$

Applications

Solving the primal via the dual

An important consequence of stationarity: under strong duality, given a dual solution λ^*, ν^*, any primal solution x^* solves

$$\min_{x \in \mathbb{R}^n} f_0(x) + \sum_{i=1}^m \lambda_i^* f_i(x) + \sum_{i=1}^p \nu_i^* h_i(x)$$

Often, solutions of this unconstrained problem can be expressed **explicitly**, giving an explicit characterization of primal solutions from dual solutions.

Furthermore, suppose the solution of this problem is unique; then it must be the primal solution x^*.

This can be very helpful when the dual is easier to solve than the primal.

Solving the primal via the dual

For example, consider:

$$
\min_x \sum_{i=1}^n f_i(x_i) \quad \text{subject to} \quad a^T x = b
$$

Solving the primal via the dual

For example, consider:

$$\min_x \sum_{i=1}^n f_i(x_i) \quad \text{subject to} \quad a^T x = b$$

where each $f_i(x_i) = \frac{1}{2} c_i x_i^2$ (smooth and strictly convex).
The dual function:

$$g(\nu) = \min_x \sum_{i=1}^n f_i(x_i) + \nu\big(b - a^T x\big)$$

Solving the primal via the dual

For example, consider:

$$\min_x \sum_{i=1}^n f_i(x_i) \quad \text{subject to} \quad a^T x = b$$

where each $f_i(x_i) = \frac{1}{2} c_i x_i^2$ (smooth and strictly convex).
The dual function:

$$\begin{aligned} g(\nu) &= \min_x \sum_{i=1}^n f_i(x_i) + \nu\left(b - a^T x\right) \\ &= b\nu + \sum_{i=1}^n \min_{x_i} \left\{ f_i(x_i) - a_i \nu x_i \right\} \end{aligned}$$

Solving the primal via the dual

For example, consider:

$$\min_x \sum_{i=1}^n f_i(x_i) \quad \text{subject to} \quad a^T x = b$$

where each $f_i(x_i) = \frac{1}{2}c_i x_i^2$ (smooth and strictly convex).
The dual function:

$$\begin{aligned} g(\nu) &= \min_x \sum_{i=1}^n f_i(x_i) + \nu\left(b - a^T x\right) \\ &= b\nu + \sum_{i=1}^n \min_{x_i} \left\{ f_i(x_i) - a_i \nu x_i \right\} \\ &= b\nu - \sum_{i=1}^n f_i^*(a_i \nu), \end{aligned}$$

Solving the primal via the dual

For example, consider:

$$\min_x \sum_{i=1}^n f_i(x_i) \quad \text{subject to} \quad a^T x = b$$

where each $f_i(x_i) = \frac{1}{2} c_i x_i^2$ (smooth and strictly convex).
The dual function:

$$\begin{aligned} g(\nu) &= \min_x \sum_{i=1}^n f_i(x_i) + \nu\left(b - a^T x\right) \\ &= b\nu + \sum_{i=1}^n \min_{x_i} \left\{ f_i(x_i) - a_i \nu x_i \right\} \\ &= b\nu - \sum_{i=1}^n f_i^*(a_i \nu), \end{aligned}$$

Solving the primal via the dual

For example, consider:

$$\min_x \sum_{i=1}^n f_i(x_i) \quad \text{subject to} \quad a^T x = b$$

where each $f_i(x_i) = \frac{1}{2} c_i x_i^2$ (smooth and strictly convex).
The dual function:

$$\begin{aligned}
g(\nu) &= \min_x \sum_{i=1}^n f_i(x_i) + \nu\left(b - a^T x\right) \\
&= b\nu + \sum_{i=1}^n \min_{x_i} \left\{ f_i(x_i) - a_i \nu x_i \right\} \\
&= b\nu - \sum_{i=1}^n f_i^*(a_i \nu),
\end{aligned}$$

where each $f_i^*(y) = \frac{1}{2c_i} y^2$, called the conjugate of f_i.

Solving the primal via the dual

For example, consider:

$$\min_x \sum_{i=1}^n f_i(x_i) \quad \text{subject to} \quad a^T x = b$$

where each $f_i(x_i) = \frac{1}{2}c_i x_i^2$ (smooth and strictly convex).
The dual function:

$$\begin{aligned} g(\nu) &= \min_x \sum_{i=1}^n f_i(x_i) + \nu\left(b - a^T x\right) \\ &= b\nu + \sum_{i=1}^n \min_{x_i} \left\{ f_i(x_i) - a_i \nu x_i \right\} \\ &= b\nu - \sum_{i=1}^n f_i^*(a_i\nu), \end{aligned}$$

where each $f_i^*(y) = \frac{1}{2c_i} y^2$, called the conjugate of f_i.

Therefore the dual problem is:

$$\max_\nu \; b\nu - \sum_{i=1}^n f_i^*(a_i\nu) \qquad \Longleftrightarrow \qquad \min_\nu \; \sum_{i=1}^n f_i^*(a_i\nu) - b\nu$$

Solving the primal via the dual

For example, consider:

$$\min_x \sum_{i=1}^n f_i(x_i) \quad \text{subject to} \quad a^T x = b$$

where each $f_i(x_i) = \frac{1}{2} c_i x_i^2$ (smooth and strictly convex). The dual function:

$$\begin{aligned} g(\nu) &= \min_x \sum_{i=1}^n f_i(x_i) + \nu \left(b - a^T x \right) \\ &= b\nu + \sum_{i=1}^n \min_{x_i} \left\{ f_i(x_i) - a_i \nu x_i \right\} \\ &= b\nu - \sum_{i=1}^n f_i^*(a_i \nu), \end{aligned}$$

where each $f_i^*(y) = \frac{1}{2c_i} y^2$, called the conjugate of f_i.

Therefore the dual problem is:

$$\max_\nu b\nu - \sum_{i=1}^n f_i^*(a_i \nu) \quad \Longleftrightarrow \quad \min_\nu \sum_{i=1}^n f_i^*(a_i \nu) - b\nu$$

This is a convex minimization problem with a scalar variable—much easier to solve than the primal.

Solving the primal via the dual

For example, consider:

$$\min_x \sum_{i=1}^n f_i(x_i) \quad \text{subject to} \quad a^T x = b$$

where each $f_i(x_i) = \frac{1}{2}c_i x_i^2$ (smooth and strictly convex). The dual function:

$$\begin{aligned} g(\nu) &= \min_x \sum_{i=1}^n f_i(x_i) + \nu\left(b - a^T x\right) \\ &= b\nu + \sum_{i=1}^n \min_{x_i} \left\{ f_i(x_i) - a_i \nu x_i \right\} \\ &= b\nu - \sum_{i=1}^n f_i^*(a_i \nu), \end{aligned}$$

where each $f_i^*(y) = \frac{1}{2c_i} y^2$, called the conjugate of f_i.

Therefore the dual problem is:

$$\max_\nu \, b\nu - \sum_{i=1}^n f_i^*(a_i\nu) \quad \Longleftrightarrow \quad \min_\nu \sum_{i=1}^n f_i^*(a_i\nu) - b\nu$$

This is a convex minimization problem with a scalar variable—much easier to solve than the primal. Given $\nu^\star$, the primal solution $x^\star$ solves:

$$\min_x \sum_{i=1}^n \left(f_i(x_i) - a_i \nu^\star x_i \right)$$

Solving the primal via the dual

For example, consider:

$$\min_x \sum_{i=1}^n f_i(x_i) \quad \text{subject to} \quad a^T x = b$$

where each $f_i(x_i) = \frac{1}{2} c_i x_i^2$ (smooth and strictly convex). The dual function:

$$\begin{aligned} g(\nu) &= \min_x \sum_{i=1}^n f_i(x_i) + \nu\left(b - a^T x\right) \\ &= b\nu + \sum_{i=1}^n \min_{x_i} \left\{ f_i(x_i) - a_i \nu x_i \right\} \\ &= b\nu - \sum_{i=1}^n f_i^*(a_i \nu), \end{aligned}$$

where each $f_i^*(y) = \frac{1}{2c_i} y^2$, called the conjugate of f_i.

Therefore the dual problem is:

$$\max_\nu \; b\nu - \sum_{i=1}^n f_i^*(a_i \nu) \quad \Longleftrightarrow \quad \min_\nu \; \sum_{i=1}^n f_i^*(a_i \nu) - b\nu$$

This is a convex minimization problem with a scalar variable—much easier to solve than the primal. Given $\nu^\star$, the primal solution $x^\star$ solves:

$$\min_x \sum_{i=1}^n \left(f_i(x_i) - a_i \nu^\star x_i \right)$$

The strict convexity of each f_i implies that this has a unique solution, namely $x^\star$, which we compute by solving $f_i'(x_i) = a_i \nu^\star$ for each i.

Solving the primal via the dual

For example, consider:

$$\min_x \sum_{i=1}^n f_i(x_i) \quad \text{subject to} \quad a^T x = b$$

where each $f_i(x_i) = \frac{1}{2}c_i x_i^2$ (smooth and strictly convex). The dual function:

$$\begin{aligned} g(\nu) &= \min_x \sum_{i=1}^n f_i(x_i) + \nu\big(b - a^T x\big) \\ &= b\nu + \sum_{i=1}^n \min_{x_i} \big\{f_i(x_i) - a_i \nu x_i\big\} \\ &= b\nu - \sum_{i=1}^n f_i^*(a_i\nu), \end{aligned}$$

where each $f_i^*(y) = \frac{1}{2c_i}y^2$, called the conjugate of f_i.

Therefore the dual problem is:

$$\max_\nu b\nu - \sum_{i=1}^n f_i^*(a_i\nu) \quad \Longleftrightarrow \quad \min_\nu \sum_{i=1}^n f_i^*(a_i\nu) - b\nu$$

This is a convex minimization problem with a scalar variable—much easier to solve than the primal.
Given $\nu^\star$, the primal solution $x^\star$ solves:

$$\min_x \sum_{i=1}^n \big(f_i(x_i) - a_i \nu^\star x_i\big)$$

The strict convexity of each f_i implies that this has a unique solution, namely $x^\star$, which we compute by solving $f_i'(x_i) = a_i\nu^\star$ for each i.
This gives:

$$x_i^\star = \frac{a_i \nu^\star}{c_i}.$$

Mixed strategies for matrix games

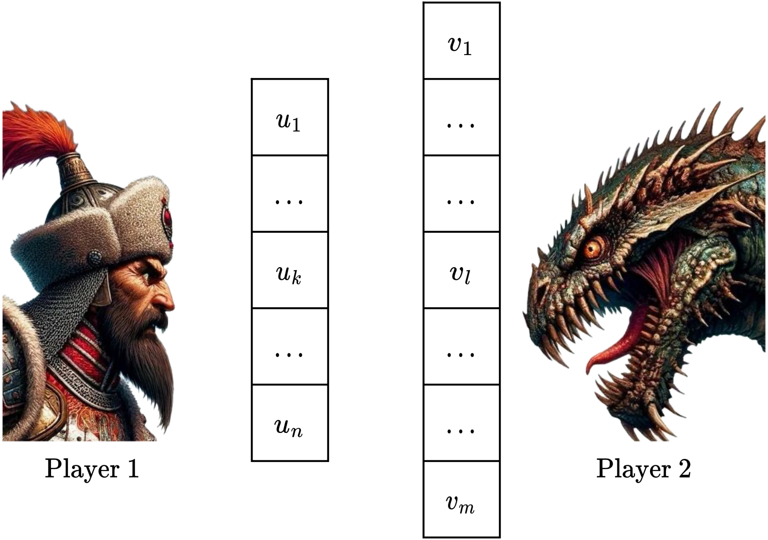

Figure 2: The scheme of a mixed strategy matrix game

Mixed strategies for matrix games

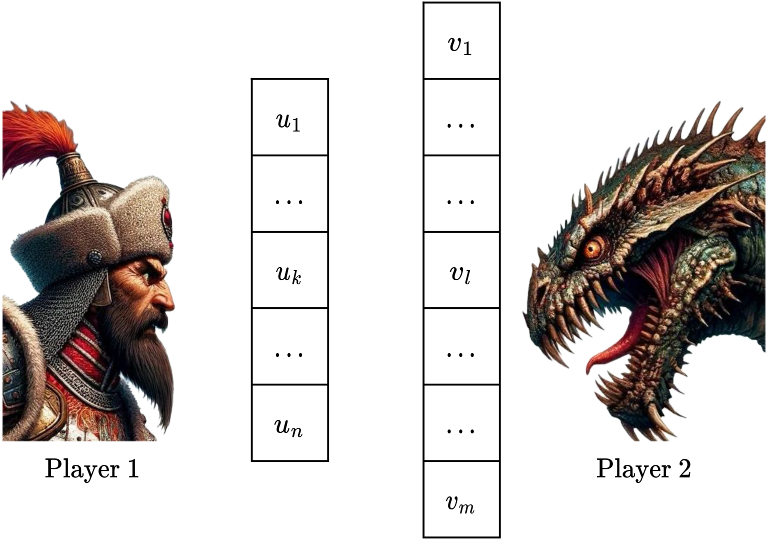

Figure 2: The scheme of a mixed strategy matrix game

In zero-sum matrix games, players 1 and 2 choose actions from sets $\{1, ..., n\}$ and $\{1, ..., m\}$, respectively. The outcome is a payment from player 1 to player 2, determined by a payoff matrix $P \in \mathbb{R}^{n \times m}$. Each player aims to use mixed strategies, choosing actions according to a probability distribution: player 1 uses probabilities u_k for each action i, and player 2 uses v_l.

Mixed strategies for matrix games

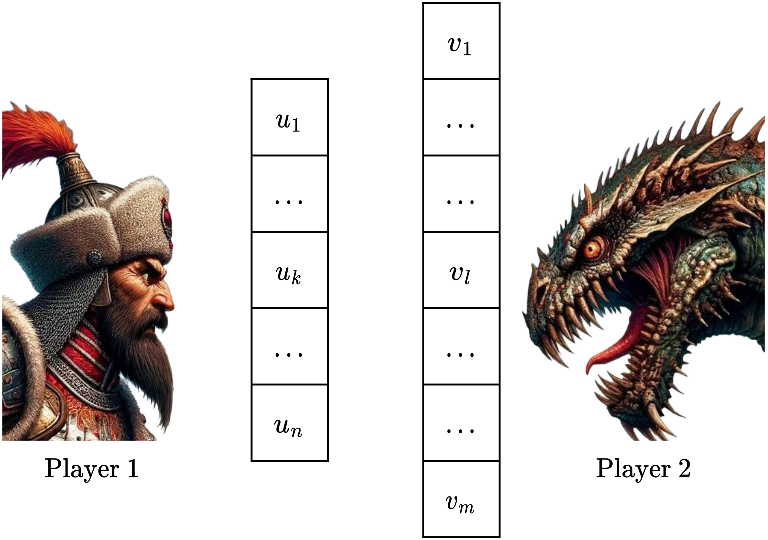

Figure 2: The scheme of a mixed strategy matrix game

In zero-sum matrix games, players 1 and 2 choose actions from sets $\{1, \dots, n\}$ and $\{1, \dots, m\}$, respectively. The outcome is a payment from player 1 to player 2, determined by a payoff matrix $P \in \mathbb{R}^{n \times m}$. Each player aims to use mixed strategies, choosing actions according to a probability distribution: player 1 uses probabilities u_k for each action i, and player 2 uses v_l.
The expected payoff from player 1 to player 2 is given by $\sum_{k=1}^{n} \sum_{l=1}^{m} u_k v_l P_{kl} = u^T P v$. Player 1 seeks to minimize this expected payoff, while player 2 aims to maximize it.

Mixed strategies for matrix games. Player 1's Perspective

Player 1

$$\begin{array}{|c|}\hline u_1 \\ \hline \ldots \\ \hline u_k \\ \hline \ldots \\ \hline u_n \\ \hline \end{array}$$

Assuming player 2 knows player 1's strategy u, player 2 will choose v to maximize $u^T P v$. The worst-case expected payoff is thus:

$$\max_{v \geq 0, 1^T v = 1} u^T P v = \max_{i=1,\ldots,m} (P^T u)_i$$

Mixed strategies for matrix games. Player 1's Perspective

Player 1

$$\begin{array}{|c|} \hline u_1 \\ \hline \dots \\ \hline u_k \\ \hline \dots \\ \hline u_n \\ \hline \end{array}$$

Assuming player 2 knows player 1's strategy u, player 2 will choose v to maximize $u^T Pv$. The worst-case expected payoff is thus:

$$\max_{v \geq 0, 1^T v = 1} u^T Pv = \max_{i=1,\dots,m} (P^T u)_i$$

Player 1's optimal strategy minimizes this worst-case payoff, leading to the optimization problem:

$$\begin{aligned} & \min \max_{i=1,\dots,m} (P^T u)_i \\ & \text{s.t. } u \geq 0 \\ & 1^T u = 1 \end{aligned} \tag{1}$$

This forms a convex optimization problem with the optimal value denoted as p_1^*.

Mixed strategies for matrix games. Player 2's Perspective

v_1
$\dots$
$\dots$
v_l
$\dots$
$\dots$
v_m

Player 2

Conversely, if player 1 knows player 2's strategy v, the goal is to minimize $u^T P v$. This leads to:

$$\min_{u \geq 0, 1^T u = 1} u^T P v = \min_{i=1,\dots,n} (Pv)_i$$

Mixed strategies for matrix games. Player 2's Perspective

$$\begin{array}{|c|}\hline v_1 \\ \hline \dots \\ \hline \dots \\ \hline v_l \\ \hline \dots \\ \hline \dots \\ \hline v_m \\ \hline \end{array}$$

Player 2

Conversely, if player 1 knows player 2's strategy v, the goal is to minimize $u^T P v$. This leads to:

$$\min_{u \geq 0, 1^T u = 1} u^T P v = \min_{i=1,\dots,n} (Pv)_i$$

Player 2 then maximizes this to get the largest guaranteed payoff, solving the optimization problem:

$$\begin{aligned} &\max \min_{i=1,\dots,n} (Pv)_i \\ &\text{s.t. } v \geq 0 \\ &1^T v = 1 \end{aligned} \tag{2}$$

The optimal value here is p_2^*.

Mixed strategies for matrix games

Duality and Equivalence

It's generally advantageous to know the opponent's strategy, but surprisingly, in mixed strategy matrix games, this advantage disappears. The key lies in duality: the problems above are Lagrange duals. By formulating player 1's problem as a linear program and introducing Lagrange multipliers, we find that the dual problem matches player 2's problem. Due to strong duality in feasible linear programs, $p_1^* = p_2^*$, showing no advantage in knowing the opponent's strategy.

Mixed strategies for matrix games

Duality and Equivalence

It's generally advantageous to know the opponent's strategy, but surprisingly, in mixed strategy matrix games, this advantage disappears. The key lies in duality: the problems above are Lagrange duals. By formulating player 1's problem as a linear program and introducing Lagrange multipliers, we find that the dual problem matches player 2's problem. Due to strong duality in feasible linear programs, $p_1^* = p_2^*$, showing no advantage in knowing the opponent's strategy.

Formulating and Solving the Lagrange Dual

We approach problem Equation 1 by setting it up as a linear programming (LP) problem. The goal is to minimize a variable t, subject to certain constraints:

Mixed strategies for matrix games

Duality and Equivalence

It's generally advantageous to know the opponent's strategy, but surprisingly, in mixed strategy matrix games, this advantage disappears. The key lies in duality: the problems above are Lagrange duals. By formulating player 1's problem as a linear program and introducing Lagrange multipliers, we find that the dual problem matches player 2's problem. Due to strong duality in feasible linear programs, $p_1^* = p_2^*$, showing no advantage in knowing the opponent's strategy.

Formulating and Solving the Lagrange Dual

We approach problem Equation 1 by setting it up as a linear programming (LP) problem. The goal is to minimize a variable t, subject to certain constraints:

1. $u \geq 0$,

Mixed strategies for matrix games

Duality and Equivalence

It's generally advantageous to know the opponent's strategy, but surprisingly, in mixed strategy matrix games, this advantage disappears. The key lies in duality: the problems above are Lagrange duals. By formulating player 1's problem as a linear program and introducing Lagrange multipliers, we find that the dual problem matches player 2's problem. Due to strong duality in feasible linear programs, $p_1^* = p_2^*$, showing no advantage in knowing the opponent's strategy.

Formulating and Solving the Lagrange Dual

We approach problem Equation 1 by setting it up as a linear programming (LP) problem. The goal is to minimize a variable t, subject to certain constraints:

1. $u \geq 0$,
2. The sum of elements in u equals 1 ($1^T u = 1$),

Mixed strategies for matrix games

Duality and Equivalence

It's generally advantageous to know the opponent's strategy, but surprisingly, in mixed strategy matrix games, this advantage disappears. The key lies in duality: the problems above are Lagrange duals. By formulating player 1's problem as a linear program and introducing Lagrange multipliers, we find that the dual problem matches player 2's problem. Due to strong duality in feasible linear programs, $p_1^* = p_2^*$, showing no advantage in knowing the opponent's strategy.

Formulating and Solving the Lagrange Dual

We approach problem Equation 1 by setting it up as a linear programming (LP) problem. The goal is to minimize a variable t, subject to certain constraints:

1. $u \geq 0$,
2. The sum of elements in u equals 1 ($1^T u = 1$),
3. $P^T u$ is less than or equal to t times a vector of ones ($P^T u \leq t\mathbf{1}$).

Mixed strategies for matrix games

Duality and Equivalence

It's generally advantageous to know the opponent's strategy, but surprisingly, in mixed strategy matrix games, this advantage disappears. The key lies in duality: the problems above are Lagrange duals. By formulating player 1's problem as a linear program and introducing Lagrange multipliers, we find that the dual problem matches player 2's problem. Due to strong duality in feasible linear programs, $p_1^* = p_2^*$, showing no advantage in knowing the opponent's strategy.

Formulating and Solving the Lagrange Dual

We approach problem Equation 1 by setting it up as a linear programming (LP) problem. The goal is to minimize a variable t, subject to certain constraints:

1. $u \geq 0$,
2. The sum of elements in u equals 1 ($1^T u = 1$),
3. $P^T u$ is less than or equal to t times a vector of ones ($P^T u \leq t\mathbf{1}$).

Mixed strategies for matrix games

Duality and Equivalence

It's generally advantageous to know the opponent's strategy, but surprisingly, in mixed strategy matrix games, this advantage disappears. The key lies in duality: the problems above are Lagrange duals. By formulating player 1's problem as a linear program and introducing Lagrange multipliers, we find that the dual problem matches player 2's problem. Due to strong duality in feasible linear programs, $p_1^* = p_2^*$, showing no advantage in knowing the opponent's strategy.

Formulating and Solving the Lagrange Dual

We approach problem Equation 1 by setting it up as a linear programming (LP) problem. The goal is to minimize a variable t, subject to certain constraints:

1. $u \geq 0$,
2. The sum of elements in u equals 1 ($1^T u = 1$),
3. $P^T u$ is less than or equal to t times a vector of ones ($P^T u \leq t\mathbf{1}$).

Here, t is an additional variable in the real numbers ($t \in \mathbb{R}$).

Mixed strategies for matrix games

Duality and Equivalence

It's generally advantageous to know the opponent's strategy, but surprisingly, in mixed strategy matrix games, this advantage disappears. The key lies in duality: the problems above are Lagrange duals. By formulating player 1's problem as a linear program and introducing Lagrange multipliers, we find that the dual problem matches player 2's problem. Due to strong duality in feasible linear programs, $p_1^* = p_2^*$, showing no advantage in knowing the opponent's strategy.

Formulating and Solving the Lagrange Dual

We approach problem Equation 1 by setting it up as a linear programming (LP) problem. The goal is to minimize a variable t, subject to certain constraints:

1. $u \geq 0$,
2. The sum of elements in u equals 1 ($1^T u = 1$),
3. $P^T u$ is less than or equal to t times a vector of ones ($P^T u \leq t\mathbf{1}$).

Here, t is an additional variable in the real numbers ($t \in \mathbb{R}$).

Constructing the Lagrangian

Mixed strategies for matrix games

Duality and Equivalence

It's generally advantageous to know the opponent's strategy, but surprisingly, in mixed strategy matrix games, this advantage disappears. The key lies in duality: the problems above are Lagrange duals. By formulating player 1's problem as a linear program and introducing Lagrange multipliers, we find that the dual problem matches player 2's problem. Due to strong duality in feasible linear programs, $p_1^* = p_2^*$, showing no advantage in knowing the opponent's strategy.

Formulating and Solving the Lagrange Dual

We approach problem Equation 1 by setting it up as a linear programming (LP) problem. The goal is to minimize a variable t, subject to certain constraints:

1. $u \geq 0$,
2. The sum of elements in u equals 1 ($1^T u = 1$),
3. $P^T u$ is less than or equal to t times a vector of ones ($P^T u \leq t\mathbf{1}$).

Here, t is an additional variable in the real numbers ($t \in \mathbb{R}$).

Constructing the Lagrangian

We introduce multipliers for the constraints: λ for $P^T u \leq t\mathbf{1}$, μ for $u \geq 0$, and ν for $1^T u = 1$. The Lagrangian is then formed as:

Mixed strategies for matrix games

Duality and Equivalence

It's generally advantageous to know the opponent's strategy, but surprisingly, in mixed strategy matrix games, this advantage disappears. The key lies in duality: the problems above are Lagrange duals. By formulating player 1's problem as a linear program and introducing Lagrange multipliers, we find that the dual problem matches player 2's problem. Due to strong duality in feasible linear programs, $p_1^* = p_2^*$, showing no advantage in knowing the opponent's strategy.

Formulating and Solving the Lagrange Dual

We approach problem Equation 1 by setting it up as a linear programming (LP) problem. The goal is to minimize a variable t, subject to certain constraints:

1. $u \geq 0$,
2. The sum of elements in u equals 1 ($1^T u = 1$),
3. $P^T u$ is less than or equal to t times a vector of ones ($P^T u \leq t\mathbf{1}$).

Here, t is an additional variable in the real numbers ($t \in \mathbb{R}$).

Constructing the Lagrangian

We introduce multipliers for the constraints: λ for $P^T u \leq t\mathbf{1}$, μ for $u \geq 0$, and ν for $1^T u = 1$. The Lagrangian is then formed as:

$$L = t + \lambda^T (P^T u - t\mathbf{1}) - \mu^T u + \nu(1 - 1^T u) = \nu + (1 - 1^T \lambda)t + (P\lambda - \nu\mathbf{1} - \mu)^T u$$

Mixed strategies for matrix games

Duality and Equivalence

It's generally advantageous to know the opponent's strategy, but surprisingly, in mixed strategy matrix games, this advantage disappears. The key lies in duality: the problems above are Lagrange duals. By formulating player 1's problem as a linear program and introducing Lagrange multipliers, we find that the dual problem matches player 2's problem. Due to strong duality in feasible linear programs, $p_1^* = p_2^*$, showing no advantage in knowing the opponent's strategy.

Formulating and Solving the Lagrange Dual

We approach problem Equation 1 by setting it up as a linear programming (LP) problem. The goal is to minimize a variable t, subject to certain constraints:

1. $u \geq 0$,
2. The sum of elements in u equals 1 ($1^T u = 1$),
3. $P^T u$ is less than or equal to t times a vector of ones ($P^T u \leq t\mathbf{1}$).

Here, t is an additional variable in the real numbers ($t \in \mathbb{R}$).

Constructing the Lagrangian

We introduce multipliers for the constraints: λ for $P^T u \leq t\mathbf{1}$, μ for $u \geq 0$, and ν for $1^T u = 1$. The Lagrangian is then formed as:

$$L = t + \lambda^T (P^T u - t\mathbf{1}) - \mu^T u + \nu(1 - 1^T u) = \nu + (1 - 1^T \lambda)t + (P\lambda - \nu\mathbf{1} - \mu)^T u$$

Mixed strategies for matrix games

Defining the Dual Function

Mixed strategies for matrix games

Defining the Dual Function

The dual function $g(\lambda, \mu, \nu)$ is defined as:

Mixed strategies for matrix games

Defining the Dual Function

The dual function $g(\lambda, \mu, \nu)$ is defined as:

$$g(\lambda, \mu, \nu) = \begin{cases} \nu & \text{if } 1^T\lambda = 1 \text{ and } P\lambda - \nu\mathbf{1} = \mu \\ -\infty & \text{otherwise} \end{cases}$$

Mixed strategies for matrix games

Defining the Dual Function

The dual function $g(\lambda, \mu, \nu)$ is defined as:

$$g(\lambda, \mu, \nu) = \begin{cases} \nu & \text{if } 1^T \lambda = 1 \text{ and } P\lambda - \nu\mathbf{1} = \mu \\ -\infty & \text{otherwise} \end{cases}$$

Solving the Dual Problem

The dual problem seeks to maximize ν under the following conditions:

1. $\lambda \geq 0$,

Mixed strategies for matrix games

Defining the Dual Function

The dual function $g(\lambda, \mu, \nu)$ is defined as:

$$g(\lambda, \mu, \nu) = \begin{cases} \nu & \text{if } 1^T\lambda = 1 \text{ and } P\lambda - \nu\mathbf{1} = \mu \\ -\infty & \text{otherwise} \end{cases}$$

Solving the Dual Problem

The dual problem seeks to maximize ν under the following conditions:

1. $\lambda \geq 0$,
2. The sum of elements in λ equals 1 ($1^T\lambda = 1$),

Mixed strategies for matrix games

Defining the Dual Function

The dual function $g(\lambda, \mu, \nu)$ is defined as:

$$g(\lambda, \mu, \nu) = \begin{cases} \nu & \text{if } 1^T\lambda = 1 \text{ and } P\lambda - \nu\mathbf{1} = \mu \\ -\infty & \text{otherwise} \end{cases}$$

Solving the Dual Problem

The dual problem seeks to maximize ν under the following conditions:

1. $\lambda \geq 0$,
2. The sum of elements in λ equals 1 ($1^T\lambda = 1$),
3. $\mu \geq 0$,

Mixed strategies for matrix games

Defining the Dual Function

The dual function $g(\lambda, \mu, \nu)$ is defined as:

$$g(\lambda, \mu, \nu) = \begin{cases} \nu & \text{if } 1^T\lambda = 1 \text{ and } P\lambda - \nu\mathbf{1} = \mu \\ -\infty & \text{otherwise} \end{cases}$$

Solving the Dual Problem

The dual problem seeks to maximize ν under the following conditions:

1. $\lambda \geq 0$,
2. The sum of elements in λ equals 1 ($1^T\lambda = 1$),
3. $\mu \geq 0$,
4. $P\lambda - \nu\mathbf{1} = \mu$.

Mixed strategies for matrix games

Defining the Dual Function

The dual function $g(\lambda, \mu, \nu)$ is defined as:

$$g(\lambda, \mu, \nu) = \begin{cases} \nu & \text{if } 1^T\lambda = 1 \text{ and } P\lambda - \nu\mathbf{1} = \mu \\ -\infty & \text{otherwise} \end{cases}$$

Solving the Dual Problem

The dual problem seeks to maximize ν under the following conditions:

1. $\lambda \geq 0$,
2. The sum of elements in λ equals 1 ($1^T\lambda = 1$),
3. $\mu \geq 0$,
4. $P\lambda - \nu\mathbf{1} = \mu$.

Mixed strategies for matrix games

Defining the Dual Function

The dual function $g(\lambda, \mu, \nu)$ is defined as:

$$g(\lambda, \mu, \nu) = \begin{cases} \nu & \text{if } 1^T\lambda = 1 \text{ and } P\lambda - \nu\mathbf{1} = \mu \\ -\infty & \text{otherwise} \end{cases}$$

Solving the Dual Problem

The dual problem seeks to maximize ν under the following conditions:

1. $\lambda \geq 0$,
2. The sum of elements in λ equals 1 ($1^T\lambda = 1$),
3. $\mu \geq 0$,
4. $P\lambda - \nu\mathbf{1} = \mu$.

Upon eliminating μ, we obtain the Lagrange dual of Equation 1:

Mixed strategies for matrix games

Defining the Dual Function

The dual function $g(\lambda, \mu, \nu)$ is defined as:

$$g(\lambda, \mu, \nu) = \begin{cases} \nu & \text{if } 1^T\lambda = 1 \text{ and } P\lambda - \nu\mathbf{1} = \mu \\ -\infty & \text{otherwise} \end{cases}$$

Solving the Dual Problem

The dual problem seeks to maximize ν under the following conditions:

1. $\lambda \geq 0$,
2. The sum of elements in λ equals 1 $(1^T\lambda = 1)$,
3. $\mu \geq 0$,
4. $P\lambda - \nu\mathbf{1} = \mu$.

Upon eliminating μ, we obtain the Lagrange dual of Equation 1:

$$\begin{aligned} &\max \nu \\ &\text{s.t. } \lambda \geq 0 \\ &1^T\lambda = 1 \\ &P\lambda \geq \nu\mathbf{1} \end{aligned}$$

Mixed strategies for matrix games

Defining the Dual Function

The dual function $g(\lambda, \mu, \nu)$ is defined as:

$$g(\lambda, \mu, \nu) = \begin{cases} \nu & \text{if } 1^T\lambda = 1 \text{ and } P\lambda - \nu\mathbf{1} = \mu \\ -\infty & \text{otherwise} \end{cases}$$

Solving the Dual Problem

The dual problem seeks to maximize ν under the following conditions:

1. $\lambda \geq 0$,
2. The sum of elements in λ equals 1 ($1^T\lambda = 1$),
3. $\mu \geq 0$,
4. $P\lambda - \nu\mathbf{1} = \mu$.

Upon eliminating μ, we obtain the Lagrange dual of Equation 1:

$$\begin{aligned} &\max \nu \\ &\text{s.t. } \lambda \geq 0 \\ &1^T\lambda = 1 \\ &P\lambda \geq \nu\mathbf{1} \end{aligned}$$

Conclusion

This formulation shows that the Lagrange dual problem is equivalent to problem Equation 2. Given the feasibility of these linear programs, strong duality holds, meaning the optimal values of Equation 1 and Equation 2 are equal.

References

- Lecture on KKT conditions (very intuitive explanation) in the course “Elements of Statistical Learning” @ KTH.

References

- Lecture on KKT conditions (very intuitive explanation) in the course "Elements of Statistical Learning" @ KTH.
- One-line proof of KKT

References

- Lecture on KKT conditions (very intuitive explanation) in the course “Elements of Statistical Learning” @ KTH.
- One-line proof of KKT
- On the Second Order Optimality Conditions for Optimization Problems with Inequality Constraints

References

* Lecture on KKT conditions (very intuitive explanation) in the course “Elements of Statistical Learning” @ KTH.
* One-line proof of KKT
* On the Second Order Optimality Conditions for Optimization Problems with Inequality Constraints
* On Second Order Optimality Conditions in Nonlinear Optimization

References

- Lecture on KKT conditions (very intuitive explanation) in the course "Elements of Statistical Learning" @ KTH.
- One-line proof of KKT
- On the Second Order Optimality Conditions for Optimization Problems with Inequality Constraints
- On Second Order Optimality Conditions in Nonlinear Optimization
- Numerical Optimization by Jorge Nocedal and Stephen J. Wright.

References

- Lecture on KKT conditions (very intuitive explanation) in the course “Elements of Statistical Learning” @ KTH.
- One-line proof of KKT
- On the Second Order Optimality Conditions for Optimization Problems with Inequality Constraints
- On Second Order Optimality Conditions in Nonlinear Optimization
- Numerical Optimization by Jorge Nocedal and Stephen J. Wright.
- Duality Uses and Correspondences lecture by Ryan Tibshirani course.