Fundamentals of Linear Algebra and Optimization
Lagrangian Duality

Jean Gallier and Jocelyn Quaintance

CIS Department
University of Pennsylvania

jean@cis.upenn.edu

March 31, 2025

Primal Minimization Problem

In this section we investigate methods to solve the *Minimization Problem* (P):

$$\begin{aligned} &\text{minimize} \quad J(v) \\ &\text{subject to} \quad \varphi_i(v) \leq 0, \quad i = 1, \dots, m. \end{aligned}$$

Primal Minimization Problem

In this section we investigate methods to solve the *Minimization Problem* (P)*:*

$$\begin{aligned} &\text{minimize} && J(v) \\ &\text{subject to} && \varphi_i(v) \leq 0, \quad i = 1, \ldots, m. \end{aligned}$$

It turns out that under certain conditions the original Problem (P), called *primal problem*, can be solved in two stages with the help another Problem (D), called the *dual problem*.

Dual Problem

The Dual Problem (D) is a *maximization problem* involving a function G, called the *Lagrangian dual*, and it is obtained by *minimizing* the *Lagrangian* $L(v, \mu)$ of Problem (P) over the variable $v \in \mathbb{R}^n$, holding μ fixed, where $L\colon \Omega \times \mathbb{R}^m_+ \to \mathbb{R}$ is given by

$$L(v, \mu) = J(v) + \sum_{i=1}^{m} \mu_i \varphi_i(v),$$

with $\mu \in \mathbb{R}^m_+$.

Duality Method for Solving Problem (P)

The two steps of the method are:

Duality Method for Solving Problem (P)

The two steps of the method are:

(1) Find the dual function $\mu \mapsto G(\mu)$ explicitly by solving the *minimization problem* of finding the minimum of $L(v, \mu)$ with respect to $v \in \Omega$, holding μ fixed.

Duality Method for Solving Problem (P)

The two steps of the method are:

(1) Find the dual function $\mu \mapsto G(\mu)$ explicitly by solving the *minimization problem* of finding the minimum of $L(v, \mu)$ with respect to $v \in \Omega$, holding μ fixed. This is an *unconstrained* minimization problem (with $v \in \Omega$). If we are lucky, a unique minimizer u_μ such that $G(\mu) = L(u_\mu, \mu)$ can be found.

Duality Method for Solving Problem (P)

The two steps of the method are:

(1) Find the dual function $\mu \mapsto G(\mu)$ explicitly by solving the *minimization problem* of finding the minimum of $L(v, \mu)$ with respect to $v \in \Omega$, holding μ fixed. This is an *unconstrained* minimization problem (with $v \in \Omega$). If we are lucky, a unique minimizer u_μ such that $G(\mu) = L(u_\mu, \mu)$ can be found.

(2) Solve the *maximization problem* of finding the maximum of the function $\mu \mapsto G(\mu)$ over all $\mu \in \mathbb{R}^m_+$.

Duality Method for Solving Problem (P)

The two steps of the method are:

(1) Find the dual function $\mu \mapsto G(\mu)$ explicitly by solving the *minimization problem* of finding the minimum of $L(v, \mu)$ with respect to $v \in \Omega$, holding μ fixed. This is an *unconstrained* minimization problem (with $v \in \Omega$). If we are lucky, a unique minimizer u_μ such that $G(\mu) = L(u_\mu, \mu)$ can be found.

(2) Solve the *maximization problem* of finding the maximum of the function $\mu \mapsto G(\mu)$ over all $\mu \in \mathbb{R}^m_+$. This is basically an unconstrained problem, except for the fact that $\mu \in \mathbb{R}^m_+$.

Duality Method for Solving Problem (P)

If Steps (1) and (2) are successful, under some suitable conditions on the function J and the constraints φ_i (for example, if they are convex), for any solution $\lambda \in \mathbb{R}_+^m$ obtained in Step (2), the vector u_λ obtained in Step (1) is an optimal solution of Problem (P).

Duality Method for Solving Problem (P)

If Steps (1) and (2) are successful, under some suitable conditions on the function J and the constraints φ_i (for example, if they are convex), for any solution $\lambda \in \mathbb{R}_+^m$ obtained in Step (2), the vector u_λ obtained in Step (1) is an optimal solution of Problem (P).

The local minima of a function $J\colon \Omega \to \mathbb{R}$ over a domain U defined by inequality constraints are *saddle points* of the Lagrangian $L(v, \mu)$ associated with J and the constraints φ_i.

Duality Method for Solving Problem (P)

If Steps (1) and (2) are successful, under some suitable conditions on the function J and the constraints φ_i (for example, if they are convex), for any solution $\lambda \in \mathbb{R}^m_+$ obtained in Step (2), the vector u_λ obtained in Step (1) is an optimal solution of Problem (P).

The local minima of a function $J \colon \Omega \to \mathbb{R}$ over a domain U defined by inequality constraints are *saddle points* of the Lagrangian $L(v, \mu)$ associated with J and the constraints φ_i.

In this presentation we do not discuss saddle points since this would take too much time.

Primal Minimization Problem

We now return to our main Minimization Problem (P):

$$\begin{aligned} &\text{minimize} && J(v) \\ &\text{subject to} && \varphi_i(v) \leq 0, \quad i = 1, \ldots, m, \end{aligned}$$

where $J \colon \Omega \to \mathbb{R}$ and the constraints $\varphi_i \colon \Omega \to \mathbb{R}$ are some functions defined on some open subset Ω of some finite-dimensional Euclidean vector space V (more generally, a real Hilbert space V).

Lagrangian of the Minimization Problem

Definition. The *Lagrangian* of the Minimization Problem (P) defined above is the function $L\colon \Omega \times \mathbb{R}^m_+ \to \mathbb{R}$ given by

$$L(v, \mu) = J(v) + \sum_{i=1}^{m} \mu_i \varphi_i(v),$$

with $\mu = (\mu_1, \ldots, \mu_m)$. The numbers μ_i are called *generalized Lagrange multipliers*.

Dual Maximization Problem

We are naturally led to introduce the function $G\colon \mathbb{R}^m_+ \to \mathbb{R}$ given by

$$G(\mu) = \inf_{v \in \Omega} L(v, \mu) \quad \mu \in \mathbb{R}^m_+,$$

and then λ will be a solution of the problem

$$\text{find } \lambda \in \mathbb{R}^m_+ \text{ such that}$$
$$G(\lambda) = \sup_{\mu \in \mathbb{R}^m_+} G(\mu),$$

Dual Maximization Problem

We are naturally led to introduce the function $G\colon \mathbb{R}^m_+ \to \mathbb{R}$ given by

$$G(\mu) = \inf_{v \in \Omega} L(v, \mu) \quad \mu \in \mathbb{R}^m_+,$$

and then λ will be a solution of the problem

$$\text{find } \lambda \in \mathbb{R}^m_+ \text{ such that}$$
$$G(\lambda) = \sup_{\mu \in \mathbb{R}^m_+} G(\mu),$$

which is equivalent to the *Maximization Problem (D):*

$$\begin{aligned} &\text{maximize} \quad G(\mu) \\ &\text{subject to} \quad \mu \in \mathbb{R}^m_+. \end{aligned}$$

Lagrangian Duality

Definition. Given the Minimization Problem (P)

$$\begin{aligned} &\text{minimize} && J(v) \\ &\text{subject to} && \varphi_i(v) \le 0, \quad i = 1, \ldots, m, \end{aligned}$$

where $J \colon \Omega \to \mathbb{R}$ and the constraints $\varphi_i \colon \Omega \to \mathbb{R}$ are some functions defined on some open subset Ω of some finite-dimensional Euclidean vector space V (more generally, a real Hilbert space V), the function $G \colon \mathbb{R}^m_+ \to \mathbb{R}$ given by

$$G(\mu) = \inf_{v \in \Omega} L(v, \mu) \quad \mu \in \mathbb{R}^m_+,$$

is called the *Lagrange dual function* (or simply *dual function*).

Lagrange Dual Problem

Problem (D)

$$\begin{aligned} \text{maximize} \quad & G(\mu) \\ \text{subject to} \quad & \mu \in \mathbb{R}^m_+ \end{aligned}$$

is called the *Lagrange dual problem*.

Lagrange Dual Problem

Problem (D)

$$\begin{aligned} \text{maximize} \quad & G(\mu) \\ \text{subject to} \quad & \mu \in \mathbb{R}^m_+ \end{aligned}$$

is called the *Lagrange dual problem*.

Problem (P) is often called the *primal problem*, and (D) is the *dual problem*. The variable μ is called the *dual variable*. The variable $\mu \in \mathbb{R}^m_+$ is said to be *dual feasible* if $G(\mu)$ is defined (not $-\infty$). If $\lambda \in \mathbb{R}^m_+$ is a maximum of G, then we call it a *dual optimal* or an *optimal Lagrange multiplier*.

Dual as a Convex Optimization Problem

Since

$$L(v, \mu) = J(v) + \sum_{i=1}^{m} \mu_i \varphi_i(v),$$

the function $G(\mu) = \inf_{v \in \Omega} L(v, \mu)$ is the pointwise infimum of some affine functions of μ, so it is *concave*, even if the φ_i are not convex.

Dual as a Convex Optimization Problem

Since

$$L(v, \mu) = J(v) + \sum_{i=1}^{m} \mu_i \varphi_i(v),$$

the function $G(\mu) = \inf_{v \in \Omega} L(v, \mu)$ is the pointwise infimum of some affine functions of μ, so it is *concave*, even if the φ_i are not convex.

One of the main advantages of the dual problem over the primal problem is that it is a *convex optimization problem*, since we wish to maximize a concave objective function G (thus minimize $-G$, a convex function), and the constraints $\mu \geq 0$ are convex. In a number of practical situations, the dual function G can indeed be computed.

Dual as a Partial Function

To be perfectly rigorous, we should mention that the dual function G is actually a *partial function*, because it takes the value $-\infty$ when the map $v \mapsto L(v, \mu)$ is unbounded below.

Dual of a Linear Program

Example. Consider the Linear Program (P)

$$\begin{aligned} \text{minimize} \quad & c^\top v \\ \text{subject to} \quad & Av \leq b, \;\; v \geq 0, \end{aligned}$$

where A is an $m \times n$ matrix, $v \in \mathbb{R}^n$, $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$.

Dual of a Linear Program

Example. Consider the Linear Program (P)

$$\begin{aligned} &\text{minimize} \quad c^\top v \\ &\text{subject to} \;\; Av \le b, \;\; v \ge 0, \end{aligned}$$

where A is an $m \times n$ matrix, $v \in \mathbb{R}^n$, $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$.

The constraints $v \ge 0$ are rewritten as $-v_i \le 0$, so we introduce Lagrange multipliers $\mu \in \mathbb{R}^m_+$ and $\nu \in \mathbb{R}^n_+$, and we have the Lagrangian

$$\begin{aligned} L(v, \mu, \nu) &= c^\top v + \mu^\top (Av - b) - \nu^\top v \\ &= -b^\top \mu + (c + A^\top \mu - \nu)^\top v. \end{aligned}$$

Dual of a Linear Program

The linear function $v \mapsto (c + A^\top \mu - \nu)^\top v$ is unbounded below unless $c + A^\top \mu - \nu = 0$, so the dual function $G(\mu, \nu) = \inf_{v \in \mathbb{R}^n} L(v, \mu, \nu)$ is given for all $\mu \geq 0$ and $\nu \geq 0$ by

$$G(\mu, \nu) = \begin{cases} -b^\top \mu & \text{if } A^\top \mu - \nu + c = 0, \\ -\infty & \text{otherwise.} \end{cases}$$

Dual of a Linear Program

The linear function $v \mapsto (c + A^\top \mu - \nu)^\top v$ is unbounded below unless $c + A^\top \mu - \nu = 0$, so the dual function $G(\mu, \nu) = \inf_{v \in \mathbb{R}^n} L(v, \mu, \nu)$ is given for all $\mu \geq 0$ and $\nu \geq 0$ by

$$G(\mu, \nu) = \begin{cases} -b^\top \mu & \text{if } A^\top \mu - \nu + c = 0, \\ -\infty & \text{otherwise.} \end{cases}$$

The domain of G is a proper subset of $\mathbb{R}^m_+ \times \mathbb{R}^n_+$.

Dual of a Linear Program

The linear function $v \mapsto (c + A^\top \mu - \nu)^\top v$ is unbounded below unless $c + A^\top \mu - \nu = 0$, so the dual function $G(\mu, \nu) = \inf_{v \in \mathbb{R}^n} L(v, \mu, \nu)$ is given for all $\mu \geq 0$ and $\nu \geq 0$ by

$$G(\mu, \nu) = \begin{cases} -b^\top \mu & \text{if } A^\top \mu - \nu + c = 0, \\ -\infty & \text{otherwise.} \end{cases}$$

The domain of G is a proper subset of $\mathbb{R}^m_+ \times \mathbb{R}^n_+$.

Observe that the value $G(\mu, \nu)$ of the function G, when it is defined, is independent of the second argument ν.

Dual of a Linear Program

Another way to obtain $G(\mu, \nu)$ is to observe that the function $v \mapsto (c + A^\top \mu - \nu)^\top v - b^\top \mu$ is affine, thus convex, and since $\mathbb{R}^n$ is convex and open, by an earlier theorem, this function (for μ, ν fixed) has a minimum at v iff

$$\nabla L(-, \mu, \nu)_v = c + A^\top \mu - \nu = 0,$$

a condition independent of v.

Dual of a Linear Program

Another way to obtain $G(\mu, \nu)$ is to observe that the function $v \mapsto (c + A^\top \mu - \nu)^\top v - b^\top \mu$ is affine, thus convex, and since $\mathbb{R}^n$ is convex and open, by an earlier theorem, this function (for μ, ν fixed) has a minimum at v iff

$$\nabla L(-, \mu, \nu)_v = c + A^\top \mu - \nu = 0,$$

a condition independent of v.

But if $c + A^\top \mu - \nu = 0$, the function $v \mapsto (c + A^\top \mu - \nu)^\top v - b^\top \mu$ is the constant function with value $-b^\top \mu$, so indeed $G(\mu, \nu)$ is defined as above.

Dual of a Linear Program

Since $G(\mu, \nu)$ is independent of ν, we introduce the function $\widehat{G}$ of the single argument μ given by

$$\widehat{G}(\mu) = -b^\top \mu,$$

which is defined for all $\mu \in \mathbb{R}^m_+$.

Dual of a Linear Program

Since $G(\mu, \nu)$ is independent of ν, we introduce the function $\widehat{G}$ of the single argument μ given by

$$\widehat{G}(\mu) = -b^\top \mu,$$

which is defined for all $\mu \in \mathbb{R}^m_+$.

Of course, $\sup_{\mu \in \mathbb{R}^m_+} \widehat{G}(\mu)$ and $\sup_{(\mu,\nu) \in \mathbb{R}^m_+ \times \mathbb{R}^n_+} G(\mu, \nu)$ are generally different, but note that $\widehat{G}(\mu) = G(\mu, \nu)$ iff there is some $\nu \in \mathbb{R}^n_+$ such that $A^\top \mu - \nu + c = 0$ iff $A^\top \mu + c \geq 0$.

Dual of a Linear Program

Since $G(\mu, \nu)$ is independent of ν, we introduce the function $\widehat{G}$ of the single argument μ given by

$$\widehat{G}(\mu) = -b^\top \mu,$$

which is defined for all $\mu \in \mathbb{R}^m_+$.

Of course, $\sup_{\mu \in \mathbb{R}^m_+} \widehat{G}(\mu)$ and $\sup_{(\mu,\nu) \in \mathbb{R}^m_+ \times \mathbb{R}^n_+} G(\mu, \nu)$ are generally different, but note that $\widehat{G}(\mu) = G(\mu, \nu)$ iff there is some $\nu \in \mathbb{R}^n_+$ such that $A^\top \mu - \nu + c = 0$ iff $A^\top \mu + c \geq 0$. Therefore, finding $\sup_{(\mu,\nu) \in \mathbb{R}^m_+ \times \mathbb{R}^n_+} G(\mu, \nu)$ is equivalent to the *constrained* Problem (D_1)

$$\begin{aligned} &\text{maximize} && -b^\top \mu \\ &\text{subject to} && A^\top \mu \geq -c, \ \ \mu \geq 0. \end{aligned}$$

Hidden Constraints Within the Dual

In summary, the dual function G of a Primary Problem (P) often contains *hidden* inequality constraints that define its domain, and sometimes it is possible to make these domain constraints $\psi_1(\mu) \leq 0, \ldots, \psi_p(\mu) \leq 0$ explicit, to define a new function $\widehat{G}$ that depends only on $q < m$ of the variables μ_i and is defined for all values $\mu_i \geq 0$ of these variables, and to replace the Maximization Problem (D), find $\sup_{\mu \in \mathbb{R}^m_+} G(\mu)$, by the constrained Problem (D_1)

$$\begin{aligned} &\text{maximize} \quad \widehat{G}(\mu) \\ &\text{subject to} \quad \psi_i(\mu) \leq 0, \quad i = 1, \ldots, p. \end{aligned}$$

Hidden Constraints Within the Dual

In summary, the dual function G of a Primary Problem (P) often contains *hidden* inequality constraints that define its domain, and sometimes it is possible to make these domain constraints $\psi_1(\mu) \leq 0, \ldots, \psi_p(\mu) \leq 0$ explicit, to define a new function $\widehat{G}$ that depends only on $q < m$ of the variables μ_i and is defined for all values $\mu_i \geq 0$ of these variables, and to replace the Maximization Problem (D), find $\sup_{\mu \in \mathbb{R}^m_+} G(\mu)$, by the constrained Problem (D_1)

$$\begin{aligned} &\text{maximize} && \widehat{G}(\mu) \\ &\text{subject to} && \psi_i(\mu) \leq 0, \quad i = 1, \ldots, p. \end{aligned}$$

Problem (D_1) is different from the Dual Program (D), but it is *equivalent* to (D) as a maximization problem.