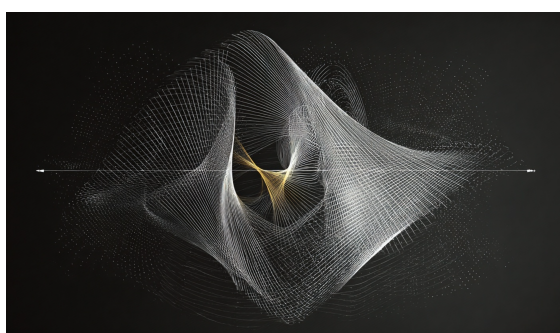


Convex Optimization and the Minimax Theorem

Stephen J. Mildenhall

2026-04-12



0.1 Lagrange Multipliers

Lagrange multipliers are often taught as a recipe: write down a Lagrangian, differentiate, solve, and then inspect the answer. That is enough to make the method usable, but it leaves several natural questions hanging. Why do the multipliers have the signs they do? Why do we optimize in opposite directions over the primal and dual variables? Why does a quantity introduced as a bookkeeping device later acquire an economic interpretation as a price?

A better way to understand the method is to begin before any formal theory and ask what we are actually trying to do. A constrained optimization problem asks us to optimize while respecting a wall. The wall may matter, or it may not. If the unconstrained optimum already lies in the feasible region, the constraint is present but inactive. If the unconstrained optimum lies outside, the constraint becomes decisive and pushes the solution onto the boundary. The Lagrangian is a device for encoding that wall in a way that can be optimized.

The goal of this post is to build that device by hand, starting with the simplest one-dimensional examples. Once the construction is clear, the familiar formal apparatus—dual variables, shadow prices, minimax, and Fenchel–Moreau style duality—stops looking like a trick and starts looking inevitable.

0.2 The Primal Problem: Optimization With a Wall

Consider first the unconstrained problem

$$\min_x x^2.$$

Its solution is $x = 0$. Nothing subtle happens: the square is always nonnegative and vanishes only at the origin.

Now add a constraint such as

$$x \leq a.$$

The feasible region is the half-line to the left of a . Two cases appear immediately. If $a \geq 0$, then 0 is feasible and the constraint changes nothing; the wall is present but irrelevant. If $a < 0$, then 0 is no longer allowed, and the best feasible point is the one closest to 0, namely $x = a$. The optimum lands on the wall.

The same picture holds for a lower bound

$$x \geq b.$$

If $b \leq 0$, the unconstrained minimizer remains feasible and the constraint does not matter. If $b > 0$, the optimum is pushed onto the boundary point $x = b$. With both bounds,

$$b \leq x \leq a,$$

the solution is again the feasible point closest to 0. If 0 lies in the interval, it is optimal; otherwise the nearest endpoint is.

That is the whole primal picture. A constraint is a wall. The only questions are whether the wall is touched, and how to encode its effect analytically.

0.3 Building the Lagrangian by Hand

Suppose we want to minimize a function $f(x)$ subject to a constraint

$$g(x) \leq 0.$$

The expression $g(x)$ is chosen so that feasible points satisfy $g(x) \leq 0$ and infeasible points satisfy $g(x) > 0$.

We now want to build a modified objective

$$L(x, \lambda) = f(x) + \lambda g(x),$$

where $\lambda \geq 0$. This is the Lagrangian.

The sign convention is not arbitrary. It is chosen so that for every feasible x ,

$$L(x, \lambda) \leq f(x).$$

Indeed, if $g(x) \leq 0$ and $\lambda \geq 0$, then $\lambda g(x) \leq 0$. So the added term does not improve any feasible point. On infeasible points, where $g(x) > 0$, the added term is positive and pushes the value upward. Thus the Lagrangian penalizes violation while leaving feasible points below the original objective. For each fixed $\lambda \geq 0$, the map $x \mapsto L(x, \lambda)$ gives a lower approximation to the constrained problem.

The Lagrangian is not yet the constrained problem itself. It is a family of penalized unconstrained problems, indexed by λ .

Remark 0.1. For a maximization problem, the sign convention in the Lagrangian reverses. If we maximize $f(x)$ subject to $g(x) \leq 0$, we write

$$L(x, \lambda) = f(x) - \lambda g(x), \quad \lambda \geq 0.$$

The reason is the same as in the minimization case: for feasible x we have $g(x) \leq 0$, so

$$-\lambda g(x) \geq 0,$$

and therefore

$$L(x, \lambda) \geq f(x).$$

Thus each fixed λ produces an upper bound on the objective over the feasible region. The dual problem then chooses the best such upper bound by minimizing over $\lambda \geq 0$.

0.4 Solving the Toy Problem

Continue with the example

$$\min_x x^2 \quad \text{subject to } x \leq a.$$

Write the constraint as

$$g(x) = x - a \leq 0.$$

The Lagrangian is

$$L(x, \lambda) = x^2 + \lambda(x - a), \quad \lambda \geq 0.$$

Now fix λ and minimize over x . Completing the square gives

$$L(x, \lambda) = \left(x + \frac{\lambda}{2}\right)^2 - \frac{\lambda^2}{4} - \lambda a.$$

So the minimizing point is

$$x(\lambda) = -\frac{\lambda}{2},$$

and the minimum value is

$$\phi(\lambda) = \inf_x L(x, \lambda) = -\frac{\lambda^2}{4} - \lambda a.$$

The function ϕ records the best lower bound produced by the particular multiplier λ . We now choose λ to make that lower bound as large as possible.

If $a \geq 0$, then the constrained optimum is still 0 at $x = 0$. In that case the best multiplier is $\lambda = 0$, and the minimizing point $x(\lambda)$ remains at the unconstrained optimum. The wall is inactive.

If $a < 0$, then the unconstrained optimum is infeasible. The minimizing point of the Lagrangian must be pushed left until it lands on the boundary:

$$-\frac{\lambda}{2} = a.$$

This gives

$$\lambda^* = -2a > 0,$$

and then

$$x^* = a.$$

The multiplier is positive precisely because the wall is active.

In both cases, the dual variable λ measures how much tilt is required to move the minimizer from the unconstrained optimum to the constrained one. When no tilt is needed, $\lambda = 0$. When the constraint binds, $\lambda > 0$.

0.5 What the Multiplier Means

Geometrically, the multiplier measures pressure from the constraint. A zero multiplier means the wall is present but exerts no force on the optimum. A positive multiplier means the wall is active and determines the answer.

At the economic level, that same quantity is naturally interpreted as a price. The primal variable describes the object being chosen—a point, a portfolio, a production plan, a level of activity. The constraint describes a scarce resource or a limit—capacity, risk tolerance, available capital, or some physical bound. The multiplier prices that limit. It is the marginal value of relaxing the constraint by a small amount. That is why the standard name shadow price is apt.

Price is not just an analogy. Later, when we place the problem inside convex duality, the multiplier really does appear as an additional dual coordinate. If a usual dual variable Q prices a payoff X through the linear pairing $Q(X)$, then adding a constraint introduces another dual coordinate λ and another quantity $g(X)$, giving the enlarged pairing

$$(Q, \lambda) \cdot (X, g(X)) = Q(X) + \lambda g(X).$$

The multiplier is therefore a price attached to the constraint quantity, just as Q is a price attached to the payoff. The only reason we call λ a shadow price rather than a market price is that it prices feasibility rather than a traded object.

That interpretation becomes fully precise once we turn to the dual problem. For now, the main lesson is simpler: the multiplier is not an arbitrary artifact of the method. It is the analytic measure of how valuable the wall is at the optimum.

0.6 The Dual Function and Why Concavity Appears

For each fixed multiplier $\lambda \geq 0$, the Lagrangian

$$L(x, \lambda) = f(x) + \lambda g(x)$$

defines an unconstrained problem in x . Its minimum value is

$$\phi(\lambda) = \inf_x L(x, \lambda).$$

ϕ is called the **dual function**. It tells us the best lower bound produced by the particular choice of λ .

The function ϕ is concave in λ because, for each fixed x , the map

$$\lambda \mapsto L(x, \lambda) = f(x) + \lambda g(x)$$

is affine in λ ; it is just a straight line. The function ϕ is obtained by taking the infimum over x of a family of straight lines. The lower envelope of affine functions is concave. Thus the dual function is concave whether or not f itself is convex.

The concavity of ϕ explains a step that otherwise looks like sleight of hand. Suppose the unconstrained maximizer of ϕ lies at a negative value of λ . Since the admissible domain is $\lambda \geq 0$, that point is unavailable. Why does the constrained maximizer then occur at $\lambda = 0$? Because a concave function cannot rise again to the right of its maximizer. Once the peak lies to the left of the feasible half-line, the function is decreasing on $[0, \infty)$, and the largest feasible value is attained at the left endpoint $\lambda = 0$.

So the familiar rule “if the candidate multiplier is negative, take $\lambda = 0$ ” is not an arbitrary truncation. It is a consequence of concavity of the dual function together with the sign restriction $\lambda \geq 0$.

0.7 Weak Duality: Every Multiplier Gives a Lower Bound

The Lagrangian was constructed so that for every feasible x and every $\lambda \geq 0$,

$$L(x, \lambda) \leq f(x).$$

Taking infimum over all x gives

$$\phi(\lambda) = \inf_x L(x, \lambda) \leq f(x)$$

for every feasible x . Hence

$$\phi(\lambda) \leq \inf_{x \in C} f(x), \quad C = \{x : g(x) \leq 0\}.$$

Since this is true for each $\lambda \geq 0$, it remains true after taking the supremum:

$$\sup_{\lambda \geq 0} \phi(\lambda) \leq \inf_{x \in C} f(x). \quad (1)$$

Equation 1 is weak duality. It says that the dual problem does not yet solve the primal problem; it merely produces lower bounds on the primal optimum. Each multiplier λ gives one such bound, and the dual problem asks for the best one.

Equation 1 is already valuable. Even when exact equality fails, the dual gives a certified lower estimate of the constrained optimum. In favorable cases, the best lower bound turns out to be exact. That exactness is the real theorem.

0.8 The Indicator Function and the Real Magic

To see what is happening it helps to fold the constraint into the objective. Define the indicator of the feasible set C by

$$I_C(x) = \begin{cases} 0, & x \in C, \\ \infty, & x \notin C. \end{cases}$$

Then the constrained problem

$$\inf_{x \in C} f(x)$$

can be written as the unconstrained problem

$$\inf_x (f(x) + I_C(x)).$$

Now the crucial identity appears:

$$I_C(x) = \sup_{\lambda \geq 0} \lambda g(x). \quad (2)$$

Indeed, if x is feasible then $g(x) \leq 0$, so the supremum over $\lambda \geq 0$ is attained at $\lambda = 0$ and equals 0. If x is infeasible then $g(x) > 0$, and $\lambda g(x) \rightarrow \infty$ as $\lambda \rightarrow \infty$, so the supremum is ∞ .

Substituting Equation 2 into the constrained objective gives

$$\begin{aligned} f(x) + I_C(x) &= f(x) + \sup_{\lambda \geq 0} \lambda g(x) \\ &= \sup_{\lambda \geq 0} (f(x) + \lambda g(x)) \\ &= \sup_{\lambda \geq 0} L(x, \lambda). \end{aligned}$$

Therefore the primal problem is

$$\inf_x \sup_{\lambda \geq 0} L(x, \lambda),$$

while the dual problem is (defined to be)

$$\sup_{\lambda \geq 0} \inf_x L(x, \lambda).$$

This is the turning point. The constraint is no longer an external side condition. It has been absorbed into the objective through a supremum over linear penalties. The multiplier λ has become part of the dual description of the problem.

0.9 Minimax: When Can We Swap the Order?

For any function $L(x, \lambda)$ we always have

$$\sup_{\lambda} \inf_x L(x, \lambda) \leq \inf_x \sup_{\lambda} L(x, \lambda),$$

which is called the minimax inequality. In our setting it is exactly weak duality, Equation 1.

Remark 0.2. To prove weak duality, compare everything to a single value $L(x, \lambda)$. Fix any $x \in X$ and any $\lambda \in \Lambda$. Since $\inf_{x'} L(x', \lambda)$ is the infimum over all x' , we have

$$\inf_{x'} L(x', \lambda) \leq L(x, \lambda).$$

Likewise, since $\sup_{\lambda'} L(x, \lambda')$ is the supremum over all λ' , we have

$$L(x, \lambda) \leq \sup_{\lambda'} L(x, \lambda').$$

Putting the two together,

$$\inf_{x'} L(x', \lambda) \leq L(x, \lambda) \leq \sup_{\lambda'} L(x, \lambda').$$

Now take supremum over λ in the left inequality. Since the right-hand side still depends on x , this gives

$$\sup_{\lambda} \inf_{x'} L(x', \lambda) \leq \sup_{\lambda} L(x, \lambda) \quad \text{for every fixed } x.$$

Because this holds for every x , we may now take infimum over x on the right:

$$\sup_{\lambda} \inf_{x'} L(x', \lambda) \leq \inf_x \sup_{\lambda} L(x, \lambda).$$

That is exactly the minimax inequality. If we minimize first, then maximize, we are optimizing a lower envelope; if we maximize first, then minimize, we are optimizing an upper envelope. A lower envelope can never exceed an upper envelope.

Example 0.1 (Function with strict minimax inequality). Minimax requires certain assumptions. Here is a simple case where it fails: a diagonal spike on $[0, 1] \times [0, 1]$:

$$L(x, \lambda) = \begin{cases} 1, & x = \lambda, \\ 0, & x \neq \lambda. \end{cases}$$

This function is 1 on the diagonal and 0 everywhere else. Fix $\lambda \in [0, 1]$, then $\inf_x L(x, \lambda) = 0$, and, since this is true for every λ , $\sup_{\lambda} \inf_x L(x, \lambda) = 0$. Now reverse the order. Fix $x \in [0, 1]$. If we choose $\lambda = x$, then $L(x, \lambda) = 1$. Hence $\sup_{\lambda} L(x, \lambda) = 1$ for every x . Taking infimum over x gives $\inf_x \sup_{\lambda} L(x, \lambda) = 1$. Thus, in this example

$$\sup_{\lambda} \inf_x L(x, \lambda) = 0 < 1 = \inf_x \sup_{\lambda} L(x, \lambda).$$

The equality fails because L is not convex in one variable, not concave in the other, and not even continuous.

The real issue is when equality holds:

$$\sup_{\lambda \geq 0} \inf_x L(x, \lambda) = \inf_x \sup_{\lambda \geq 0} L(x, \lambda).$$

When does hold the dual lower bound is exact and the dual problem recovers the primal optimum. This is called **strong duality**.

Why should such an interchange be valid? Broadly, because the Lagrangian has the right saddle structure. For each fixed λ , the map $x \mapsto L(x, \lambda)$ should be convex, so minimization in x is well behaved. For each fixed x , the map $\lambda \mapsto L(x, \lambda)$ is affine, hence concave, so maximization in λ is equally well behaved. Under enough closedness and interiority—for instance, convex objective, convex constraints, and a strictly feasible point—a minimax theorem applies and the order of inf and sup may be swapped. See Section 0.19.

0.10 Primal Optimal, Dual Optimal, Minimax, and Complementary Slackness

with some duplication, but to hammer the point home

Consider the standard **minimization problem**

$$\min_x f(x) \quad \text{subject to} \quad g_i(x) \leq 0, \quad i = 1, \dots, m.$$

Define the feasible set

$$C = \{x : g_i(x) \leq 0 \text{ for all } i\},$$

and the Lagrangian

$$L(x, \lambda) = f(x) + \sum_i \lambda_i g_i(x), \quad \lambda_i \geq 0.$$

A point x^* is **primal feasible** if it satisfies all the constraints:

$$g_i(x^*) \leq 0 \quad \text{for all } i.$$

It is **primal optimal** if it attains the smallest feasible value:

$$f(x^*) = \inf_{x \in C} f(x).$$

So the primal problem is the original constrained optimization problem in the actual decision variable x .

For each fixed $\lambda \geq 0$, define the **dual function**

$$\phi(\lambda) = \inf_x L(x, \lambda).$$

Because $L(x, \lambda) \leq f(x)$ for feasible x , each $\phi(\lambda)$ is a lower bound on the primal optimum. The **dual problem** is therefore

$$\sup_{\lambda \geq 0} \phi(\lambda) = \sup_{\lambda \geq 0} \inf_x L(x, \lambda).$$

A vector λ^* is **dual feasible** if $\lambda_i^* \geq 0$ for all i . It is **dual optimal** if it attains the largest dual value:

$$\phi(\lambda^*) = \sup_{\lambda \geq 0} \phi(\lambda).$$

The dual problem is not choosing the portfolio or production plan or weights. It is choosing the best multiplier vector, that is, the best lower bound.

We always have **weak duality**

$$\sup_{\lambda \geq 0} \inf_x L(x, \lambda) \leq \inf_{x \in C} f(x) :$$

the best dual lower bound is still no larger than the primal optimum.

Using the indicator of the feasible set,

$$I_C(x) = \begin{cases} 0, & x \in C, \\ \infty, & x \notin C, \end{cases}$$

the primal problem can be written

$$\inf_x \sup_{\lambda \geq 0} L(x, \lambda),$$

while the dual problem is

$$\sup_{\lambda \geq 0} \inf_x L(x, \lambda).$$

In general

$$\sup \inf \leq \inf \sup .$$

If equality holds,

$$\sup_{\lambda \geq 0} \inf_x L(x, \lambda) = \inf_x \sup_{\lambda \geq 0} L(x, \lambda),$$

then there is **no duality gap**. This is **strong duality**, or the relevant minimax equality. Under convexity plus a constraint qualification, this equality holds.

Suppose now that:

- x^* is primal optimal,
- λ^* is dual optimal,
- strong duality holds.

Then

$$f(x^*) = \inf_{x \in C} f(x) = \sup_{\lambda \geq 0} \inf_x L(x, \lambda) = \inf_x L(x, \lambda^*).$$

Since x^* is feasible,

$$L(x^*, \lambda^*) = f(x^*) + \sum_i \lambda_i^* g_i(x^*).$$

But feasibility gives $g_i(x^*) \leq 0$, and dual feasibility gives $\lambda_i^* \geq 0$, so

$$L(x^*, \lambda^*) \leq f(x^*).$$

On the other hand, because x^* minimizes $L(\cdot, \lambda^*)$,

$$\inf_x L(x, \lambda^*) \leq L(x^*, \lambda^*).$$

Putting everything together,

$$f(x^*) = \inf_x L(x, \lambda^*) \leq L(x^*, \lambda^*) \leq f(x^*).$$

Therefore all inequalities are equalities, so

$$L(x^*, \lambda^*) = f(x^*).$$

Now expand that equality:

$$f(x^*) + \sum_i \lambda_i^* g_i(x^*) = f(x^*).$$

Hence

$$\sum_i \lambda_i^* g_i(x^*) = 0.$$

But each term satisfies

$$\lambda_i^* g_i(x^*) \leq 0,$$

because $\lambda_i^* \geq 0$ and $g_i(x^*) \leq 0$. A sum of nonpositive terms can equal 0 only if each term vanishes:

$$\lambda_i^* g_i(x^*) = 0 \quad \text{for each } i,$$

which is **complementary slackness**. It means:

- either $g_i(x^*) < 0$, so the constraint is slack, and then necessarily $\lambda_i^* = 0$;
- or $\lambda_i^* > 0$, and then necessarily $g_i(x^*) = 0$, so the constraint binds.

So minimax gives equality of primal and dual values, and that equality forces complementary slackness.

For a **maximization problem** with constraints $g_i(x) \leq 0$, the Lagrangian is written

$$L(x, \lambda) = f(x) - \sum_i \lambda_i g_i(x), \quad \lambda_i \geq 0.$$

Then each fixed λ gives an **upper bound** on the feasible objective, the **dual problem** minimizes over λ , and **strong duality** again means max-min equals min-max in the appropriate form. **Complementary slackness** still reads

$$\lambda_i^* g_i(x^*) = 0.$$

The dictionary is:

- primal optimal = best feasible choice of the real decision variables;
- dual optimal = best feasible choice of the multipliers;
- minimax / strong duality = primal value equals dual value;
- complementary slackness = at optimality, a constraint is either inactive or priced, but not both.

0.11 Equality constraints

So far the discussion has focused on one-sided constraints such as

$$x \leq a.$$

The equality constraint

$$x = a$$

fits exactly the same framework, except that the multiplier is no longer restricted to be nonnegative. Let

$$I_{\{a\}}(x) = \begin{cases} 0, & x = a, \\ \infty, & x \neq a. \end{cases}$$

be the indicator of the singleton set $\{a\}$. Then

$$I_{\{a\}}(x) = \sup_{\mu \in \mathbb{R}} \mu(x - a).$$

Thus a constrained problem such as

$$\min_x f(x) \quad \text{subject to} \quad x = a$$

can be written as

$$\inf_x (f(x) + I_{\{a\}}(x)) = \inf_x \sup_{\mu \in \mathbb{R}} (f(x) + \mu(x - a)).$$

The Lagrangian is therefore

$$L(x, \mu) = f(x) + \mu(x - a), \quad \mu \in \mathbb{R}.$$

The associated dual function is

$$\phi(\mu) = \inf_x L(x, \mu) = \inf_x (f(x) + \mu(x - a)),$$

and the dual problem is

$$\sup_{\mu \in \mathbb{R}} \phi(\mu).$$

This mirrors the one-sided case exactly. For the inequality $x \leq a$, the multiplier ranges over $[0, \infty)$ because the indicator of a half-line is obtained by a supremum over nonnegative slopes. For the equality $x = a$, the multiplier ranges over all of \mathbb{R} because the indicator of a point requires lines of both positive and negative slope. That is the clean reason equality constraints have one unrestricted multiplier rather than two nonnegative ones.

If minimax applies, then

$$\sup_{\mu \in \mathbb{R}} \inf_x L(x, \mu) = \inf_x \sup_{\mu \in \mathbb{R}} L(x, \mu) = \inf_x (f(x) + I_{\{a\}}(x)) = f(a).$$

So the same pattern reappears: the dual variable μ prices the constraint quantity $x - a$, and strong duality says the best such pricing recovers the constrained primal value exactly.

0.12 The KKT conditions

The Karush–Kuhn–Tucker conditions package the information carried by primal feasibility, dual feasibility, stationarity, and complementary slackness. For the convex minimization problem

$$\min_x f(x) \quad \text{subject to} \quad g_i(x) \leq 0, \quad i = 1, \dots, m, \quad h_j(x) = 0, \quad j = 1, \dots, p,$$

the Lagrangian is

$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^m \lambda_i g_i(x) + \sum_{j=1}^p \mu_j h_j(x),$$

with

$$\lambda_i \geq 0, \quad \mu_j \in \mathbb{R}.$$

The KKT conditions at a candidate optimum x^* are:

1. **Primal feasibility:**

$$g_i(x^*) \leq 0, \quad h_j(x^*) = 0.$$

2. **Dual feasibility:**

$$\lambda_i^* \geq 0.$$

3. **Complementary slackness:**

$$\lambda_i^* g_i(x^*) = 0 \quad \text{for each } i.$$

This says that each inequality constraint is in one of three states. It may be **slack**, meaning

$$g_i(x^*) < 0, \quad \lambda_i^* = 0.$$

It may be **active** with positive value,

$$g_i(x^*) = 0, \quad \lambda_i^* > 0.$$

Or it may be active with zero value,

$$g_i(x^*) = 0, \quad \lambda_i^* = 0.$$

The last case is often called **weakly active**: the boundary is touched, but the constraint has no marginal value at that optimum.

4. **Stationarity** balances the objective against the priced constraints. In the differentiable case,

$$\nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla g_i(x^*) + \sum_{j=1}^p \mu_j^* \nabla h_j(x^*) = 0.$$

In the convex but possibly nonsmooth case, the same condition is written with subdifferentials:

$$0 \in \partial f(x^*) + \sum_{i=1}^m \lambda_i^* \partial g_i(x^*) + \sum_{j=1}^p \mu_j^* \partial h_j(x^*).$$

When the equality constraints are smooth or affine, one often keeps them in gradient form and writes

$$0 \in \partial f(x^*) + \sum_{i=1}^m \lambda_i^* \partial g_i(x^*) + \sum_{j=1}^p \mu_j^* \nabla h_j(x^*).$$

Under convexity together with a suitable constraint qualification, these conditions are not merely necessary; they are sufficient for optimality. In that setting, solving the problem means finding a point and multipliers that satisfy the full KKT system. The multipliers then identify which constraints truly matter, how much they matter, and how the objective gradient is balanced by the geometry of the active constraints.

0.13 Comparison with Fenchel–Moreau

What we have described so far sits naturally inside convex duality. The true primal function is not f alone but

$$F(x) = f(x) + I_C(x).$$

Under the usual convexity and lower-semicontinuity assumptions, F is a proper convex lower-semicontinuous function. Fenchel–Moreau tells us that such a function can be recovered from its affine minorants:

$$F(x) = \sup_y \langle y, x \rangle - F^*(y).$$

The Lagrangian construction is a special case of the same architecture. The indicator I_C is represented as a supremum over linear penalty terms, so the constrained problem becomes

$$F(x) = \sup_{\lambda \geq 0} L(x, \lambda).$$

Thus both Fenchel–Moreau and Lagrange duality follow the same pattern: construct a family of affine or linear lower approximations, then recover the primal object as the supremum of that family.

Weak duality mirrors the basic fact that a biconjugate lies below the original function. Strong duality mirrors exact recovery: under convexity, lower semicontinuity, and the relevant regularity assumptions, the lower approximations are rich enough to reconstruct the primal value without loss.

Seen in this way, the multiplier is no longer a strange extra variable introduced by a trick. It is simply a dual coordinate. Adding a constraint enlarges the dual side of the problem, and the multiplier is the price attached to that new coordinate.

0.14 The Price-Cost-Margin Model

The simplest economic interpretation begins with a production or trading problem. Let x denote a vector of activity levels or portfolio volumes, let q be a market price vector, and let $\alpha(x)$ be the cost of manufacturing the position x . The unconstrained margin is

$$m(q) = \sup_x qx - \alpha(x).$$

It is the best net gain attainable at price system q , with the technology α (economist's short-run).

Now impose an operating constraint and corresponding feasible region

$$g(x) \leq 0, \quad C = \{x : g(x) \leq 0\}.$$

The constrained margin becomes

$$\begin{aligned} m_C(q) &= \sup_{x \in C} qx - \alpha(x) \\ &= \sup_x qx - \alpha(x) - I_C(x). \end{aligned}$$

It represents the best net gain attainable under q including the effect of the constraints. Using the indicator representation

$$I_C(x) = \sup_{\lambda \geq 0} \lambda g(x),$$

we obtain

$$m_C(q) = \sup_x \inf_{\lambda \geq 0} qx - \alpha(x) - \lambda g(x).$$

If minimax applies, we may swap the order and write as:

$$m_C(q) = \inf_{\lambda \geq 0} \sup_x qx - \lambda g(x) - \alpha(x). \quad (3)$$

Equation 3 helps give the multiplier an economic meaning. The term

$$qx - \lambda g(x)$$

is an adjusted revenue. The vector q prices the actual activity x ; the scalar λ prices use of the constrained resource through $g(x)$. Since λ attaches a price to a limit rather than a traded object, it is naturally interpreted as a shadow price. Notice that λ enters linearly, the same as a price.

Remark 0.3 (Complementary slackness and the shadow-price interpretation). Consider the maximization problem

$$\max_x qx - f(x) \quad \text{subject to} \quad x \leq a.$$

Its Lagrangian is

$$L(x, \lambda) = qx - f(x) - \lambda(x - a), \quad \lambda \geq 0.$$

Rearranging gives

$$L(x, \lambda) = qx + \lambda a - (f(x) + \lambda x).$$

This form allows an economic interpretation. The term qx is sales revenue. The term $f(x)$ is production cost. The extra terms involving λ are notional rather than actual cash flows: λa is the imputed value of having a units of the scarce resource available, and λx is the imputed cost of using x units of that resource in this project. Thus the project is evaluated as if it owned capacity worth λ per unit and had to pay the same λ per unit for each unit consumed.

At an optimum (x^*, λ^*) , we have the condition

$$\lambda^*(x^* - a) = 0.$$

This is called **complementary slackness**. It says that at least one of the two terms must vanish. There are two cases. Either the constraint does not bind, in which case $x^* < a$ and therefore $\lambda^* = 0$ (remember, inf over $\lambda \geq 0$); or the constraint binds, in which case $x^* = a$ and λ^* may be positive. In either case the product vanishes.

Complementary slackness explains why the shadow-price term does not alter the realized objective value at the optimum. If the constraint is slack then the shadow price is zero; if the constraint binds then the resource usage equals the available amount, so the imputed value of the endowment λa exactly offsets the imputed usage cost λx . Thus no actual penalty is paid at the chosen optimum. The role of the shadow-price term is not to change the attained value after the fact, but to steer the optimization toward the correct choice.

The multiplier λ is the shadow price of the scarce resource. It measures the marginal value of relaxing the constraint by one unit, or equivalently the marginal contribution to optimal profit of increasing a slightly. If a can be purchased in a market, then λ is the bid price the project is willing to pay for one more unit. If several disjoint projects compete for the same scarce resource, then their respective shadow prices tell us where that resource is most valuable at the margin. A higher shadow price means that an extra unit of capacity is more productive in that use. These considerations apply only at the optimum.

Remember, λ is not an actual payment: it is an imputed or opportunity cost. The constraint acts as though the resource had price λ , and that “as though” is exactly what guides allocation, pricing, and comparison across competing uses.

Remark 0.4 (Conditions for complementary slackness). Complementary slackness is not automatic for an arbitrary constrained problem. It is an optimality condition that appears when a Lagrange multiplier description is valid. Weak duality always gives

$$\sup_{\lambda} \inf_x L(x, \lambda) \leq \inf_x \sup_{\lambda} L(x, \lambda)$$

and complementary slackness appears when primal and dual optimizers exist and the duality gap is zero.

For a minimization problem with constraints $g_i(x) \leq 0$ and Lagrangian

$$L(x, \lambda) = f(x) + \sum_i \lambda_i g_i(x), \quad \lambda_i \geq 0,$$

suppose x^* is primal optimal and λ^* is dual optimal, and suppose strong duality holds:

$$f(x^*) = \inf_{x \in C} f(x) = \sup_{\lambda \geq 0} \inf_x L(x, \lambda).$$

Then

$$f(x^*) = L(x^*, \lambda^*) = f(x^*) + \sum_i \lambda_i^* g_i(x^*).$$

Hence

$$\sum_i \lambda_i^* g_i(x^*) = 0.$$

But each term satisfies

$$\lambda_i^* \geq 0, \quad g_i(x^*) \leq 0,$$

so each product is ≤ 0 . A sum of nonpositive terms can be 0 only if each term is 0. Therefore

$$\lambda_i^* g_i(x^*) = 0 \quad \text{for every } i.$$

There is a very close connection with minimax and strong duality: if you have primal and dual optimizers and max-min equals min-max, then complementary slackness follows immediately; but conversely, complementary slackness by itself is not enough to give strong duality.

0.15 A Portfolio Problem With a Monetary Risk Measure

Now consider an investor who chooses portfolio weights w . Let Π be a market pricing vector and P an objective value vector. The investor wants to **maximize** expected gain relative to objective value,

$$(\Pi - P) \cdot w,$$

subject to a limit on risk-bearing capacity. If $X(w)$ is the loss associated with portfolio w and ρ is the investor's monetary risk measure, the problem is

$$\max_w (\Pi - P) \cdot w \quad \text{subject to} \quad \rho(X(w)) \leq K,$$

where K is the available risk-bearing capacity. K is usually capital, but it could combine actual balance sheet capital with contingent commitments, so we prefer the broader term.

The Lagrangian is

$$L(w, \lambda) = (\Pi - P) \cdot w - \lambda(\rho(X(w)) - K), \quad \lambda \geq 0.$$

Again λ is the shadow price of the constraint. It measures the marginal value of an extra unit of risk-bearing capacity.

If the optimum occurs at an interior point of the feasible risk region, then $\lambda = 0$ and the risk constraint is inactive. The interesting case is when the constraint binds:

$$\rho(X(w^*)) = K, \quad \lambda^* > 0.$$

Then the portfolio is chosen by balancing reward against marginal risk cost.

If $\rho \circ X$ is differentiable, the first-order condition at an optimum is

$$\Pi - P = \lambda \nabla(\rho \circ X)(w^*).$$

Thus excess reward must be proportional to marginal risk contribution. If the investor wants a particular target portfolio, say

$$w^* = \mathbf{1},$$

to be optimal, then they choose Π so that

$$\Pi = P + \lambda \nabla(\rho \circ X)(\mathbf{1}),$$

together with the binding condition

$$\rho(X(\mathbf{1})) = K.$$

In subgradient form the condition is

$$\Pi - P \in \lambda \partial(\rho \circ X)(\mathbf{1}).$$

This is an inverse-optimization view of pricing. One chooses the market pricing vector so that the desired portfolio is exactly the one justified by the investor's risk preferences and available risk capacity. The multiplier λ is again a price, now the price of relaxing the risk constraint.

Example 0.2 (Spectral ρ). Suppose the portfolio loss is linear in the weights,

$$X(w) = \sum_{i=1}^n w_i X_i,$$

interpreting w_i as a participation percentage, and suppose the risk measure ρ is spectral, associated with a distortion g . At points where ρ is differentiable, the marginal risk contribution of unit i is

$$\frac{\partial}{\partial w_i} \rho(X(w)) = P(X_i g'(S_{X(w)}(X(w))), \quad (4)$$

where $S_Y(y) = 1 - F_Y(y)$ denotes the survival function of a random variable Y . Equation 4 is the natural allocation.

In particular, at the target portfolio $w = \mathbf{1}$ (the market portfolio), with aggregate (market) loss

$$M = \sum_{j=1}^n X_j,$$

we have

$$\left. \frac{\partial}{\partial w_i} \rho(X(w)) \right|_{w=\mathbf{1}} = \mathbb{P}(X_i g'(S_M(M))).$$

Therefore, in the constrained optimization problem

$$\max_w (\Pi - \mathbb{P}) \cdot w \quad \text{subject to} \quad \rho(X(w)) \leq K,$$

the first-order condition at $w = \mathbf{1}$ becomes

$$\Pi_i - \mathbb{P}_i = \lambda \mathbb{P}(X_i g'(S_M(M))), \quad i = 1, \dots, n.$$

Thus the excess pricing vector $\Pi - \mathbb{P}$ is proportional to the vector of marginal spectral risk contributions at the target portfolio.

This formula is the spectral analogue of the general condition

$$\Pi - \mathbb{P} = \lambda \nabla(\rho \circ X)(\mathbf{1}),$$

and shows concretely how market reward must align with marginal cost of risk if the portfolio $w = \mathbf{1}$ is to be optimal.

0.16 The Dual Representation and Enlarging the Dual Space

Suppose now the monetary risk measure itself has a dual representation

$$\rho(X) = \sup_{\mathbb{Q}} \mathbb{Q}(X) - \alpha(\mathbb{Q}),$$

where \mathbb{Q} ranges over dual pricing or stress measures and $\alpha(\mathbb{Q})$ is the associated penalty. In this representation, \mathbb{Q} is already a price-like object: it assigns linear value to the payoff or loss X , while α records the cost of using that pricing rule. When ρ is coherent the penalty α becomes a convex indicator function taking the value 0 on \mathbb{Q} deemed possible and ∞ on those deemed impossible. Then $\rho(X)$ is represented as the worst scenario outcome: $\rho = \sup_{\mathbb{Q}} \mathbb{Q}(X)$.

Now add the side constraint

$$g(X) = \rho(X) - K \leq 0.$$

Then the constrained objective is

$$\rho(X) + I_C(X), \quad C = \{X : g(X) \leq 0\}.$$

Using the indicator representation,

$$I_C(X) = \sup_{\lambda \geq 0} \lambda g(X),$$

we obtain

$$\rho(X) + I_C(X) = \sup_{\mathbb{Q}} \{\mathbb{Q}(X) - \alpha(\mathbb{Q})\} + \sup_{\lambda \geq 0} \lambda g(X).$$

Combining the two suprema gives

$$\rho(X) + I_C(X) = \sup_{\mathbf{Q}, \lambda \geq 0} \mathbf{Q}(X) + \lambda g(X) - \alpha(\mathbf{Q}).$$

The dual space has expanded from \mathbf{Q} to (\mathbf{Q}, λ) , but the pairing remains linear:

$$(\mathbf{Q}, \lambda) \cdot (X, g(X)) = \mathbf{Q}(X) + \lambda g(X).$$

Thus λ enters in exactly the same way that \mathbf{Q} enters. The former prices the constraint quantity $g(X)$, while the latter prices the payoff or loss X . The multiplier is therefore a price in the same formal sense as \mathbf{Q} ; it is simply attached to a different object.

This observation unifies the economic interpretation. In the price-cost-margin model, λ prices a scarce operating resource. In the risk-measure model, λ prices risk-bearing capacity or feasibility. In both cases the dual variable is enlarged by adding one more linear pricing coordinate. That is why the phrase shadow price is more than suggestive language. It describes exactly what the multiplier is doing in the dual representation.

0.17 Adding Participation Constraints and a Cost of Capital

Consider an intermediary who chooses participation levels w_i in deals $i = 1, \dots, n$, with

$$0 \leq w_i \leq 1.$$

Let Π_i denote the market price of deal i , and let P_i denote its expected loss. The intermediary raises capital K at unit cost r , and requires capital sufficient to support the portfolio risk:

$$\rho(X(w)) \leq K.$$

Assume

$$X(w) = \sum_{i=1}^n w_i X_i,$$

and that the capital risk measure is coherent with dual representation

$$\rho(X) = \sup_{\mathbf{Q} \in \mathcal{Q}} \mathbf{Q}(X).$$

The intermediary's problem is

$$\max_{w, K} \langle \Pi - P, w \rangle - rK \quad \text{subject to} \quad \rho(X(w)) \leq K, \quad w_i \leq 1.$$

We assume dominated deals have already been removed, so lower bounds $w_i \geq 0$ do not bind and can be ignored.

Introduce the multiplier $\eta \geq 0$ for the capital constraint and multipliers $\lambda_i \geq 0$ for the upper bounds $w_i \leq 1$. The Lagrangian is

$$L(w, K, \eta, \lambda) = \langle \Pi - P, w \rangle - rK - \eta(\rho(X(w)) - K) - \sum_i \lambda_i(w_i - 1).$$

Expanding and re-writing to draw out the symmetry gives

$$L(w, K, \eta, \lambda) = \langle (\Pi - P, \eta - r), (w, K) \rangle - \eta \rho(X(w)) - \sum_i \lambda_i(w_i - 1).$$

Now optimize over the primal variables. The coefficient of K is $\eta - r$. Since K is a free choice variable, first-order optimality forces

$$\eta = r.$$

Thus the shadow price of the capital constraint equals the actual market cost of capital.

Substituting $\eta = r$ reduces the problem to

$$\max_w \langle \Pi - P, w \rangle - r \rho(X(w)) \quad \text{subject to} \quad w_i \leq 1.$$

If $\rho \circ X$ is differentiable, the first-order condition for w_i is

$$\Pi_i - P_i - r \frac{\partial}{\partial w_i} \rho(X(w)) - \lambda_i = 0.$$

Hence

$$\lambda_i = \Pi_i - P_i - r \frac{\partial}{\partial w_i} \rho(X(w)).$$

Together with complementary slackness,

$$\lambda_i(w_i - 1) = 0,$$

this gives the familiar interpretation:

- If $w_i < 1$, then $\lambda_i = 0$ and

$$\Pi_i - P_i = r \frac{\partial}{\partial w_i} \rho(X(w)).$$

- If $w_i = 1$, then

$$\Pi_i - P_i \geq r \frac{\partial}{\partial w_i} \rho(X(w)),$$

with strict inequality corresponding to positive shadow value of the volume cap.

Next, we use the coherent representation of ρ . At a point $S = X(w)$, if the supremum is attained by some supporting measure \mathbf{Q}^* , then for linear portfolios the marginal risk contribution of deal i is

$$\frac{\partial}{\partial w_i} \rho(X(w)) = \mathbf{Q}^*(X_i),$$

or, more generally,

$$\mathbf{Q}^*(X_i) \in \partial_{w_i}(\rho \circ X)(w).$$

The first-order condition becomes

$$\lambda_i = \Pi_i - P_i - r \mathbf{Q}^*(X_i).$$

This expression decomposes pricing into:

- P_i is expected loss;
- $r \mathbf{Q}^*(X_i)$ is the capitalized marginal risk charge;
- λ_i is the residual shadow value of the participation cap.

We can determine prices Π_i that support market clearing, meaning that full participation on each deal $w^* = \mathbf{1}$ is optimal. In that case, each upper-bound constraint binds, so complementary slackness allows $\lambda_i \geq 0$, and the supporting-price condition is

$$\Pi_i - P_i - r \mathbf{Q}^*(X_i) = \lambda_i \geq 0.$$

Equivalently,

$$\Pi_i \geq P_i + r \mathbf{Q}^*(X_i).$$

For all deals to be written at 100 percent by a single intermediary, each market price must again cover the expected loss, the marginal cost of capital, and possibly a residual scarcity rent λ_i if the cap $w_i \leq 1$ is binding. If we also impose competitive market-clearing then standard equilibrium condition is that these residual rents are competed away. In that case all $\lambda_i = 0$ and prices are bid down to

$$\Pi_i = P_i + r \mathbf{Q}^*(X_i).$$

Hence we arrive at the analogue of pricing at marginal cost. The market price equals expected loss plus the cost of the marginal capital required to support the deal inside the aggregate portfolio.

0.18 TL;DR

The Lagrangian is best understood as a way to replace a hard constraint by a priced penalty. Starting from a primal problem with a wall, we build

$$L(x, \lambda) = f(x) + \lambda g(x)$$

in the minimization case, or the corresponding sign-adjusted form in the maximization case, so that feasible points are never artificially improved. For each fixed multiplier, the Lagrangian gives a lower approximation to the constrained problem. The dual problem then asks for the best such lower bound.

The multiplier is not merely a technical device. When the constraint is inactive, its value is 0; when the constraint binds, it becomes positive and measures the marginal value of relaxing the constraint. In that sense it is naturally a price. More precisely, it is a shadow price: a price attached not to a traded payoff but to a scarce resource, a feasibility requirement, or a unit of risk-bearing capacity.

The deeper structure appears when the constraint is absorbed into the objective through the indicator function

$$I_C(x) = \begin{cases} 0, & x \in C, \\ \infty, & x \notin C. \end{cases}$$

Because that indicator can be written as a supremum over linear penalties, the primal problem becomes

$$\inf_x \sup_{\lambda} L(x, \lambda),$$

while the dual becomes

$$\sup_{\lambda} \inf_x L(x, \lambda).$$

Weak duality is automatic; strong duality is the statement that the two orders of optimization agree. That agreement is the real theorem, and it rests on convexity, concavity, closedness, and interiority.

From the convex-duality point of view, the multiplier simply enlarges the dual space. If a dual variable Q already prices a payoff X through the linear pairing $Q(X)$, then adding a constraint introduces an additional coordinate λ and an additional quantity $g(X)$, giving

$$(Q, \lambda) \cdot (X, g(X)) = Q(X) + \lambda g(X).$$

The dual space expands from Q to (Q, λ) , but the pairing remains linear. That is the cleanest explanation of why the multiplier deserves to be called a price.

Fenchel–Moreau and Lagrange duality therefore belong to the same story. Both recover a primal object from a family of affine or linearized lower approximants. Both rest on the same convex-analytic machinery. The multiplier enters that story not as an ad hoc extra variable, but as the new dual coordinate required to price the constraint.

0.19 Appendix: Proof of Minimax

The minimax theorem says: if L is **convex** in the **minimization** variable, **concave** in the **maximization** variable, and we have enough closedness plus some compactness/interiority to rule out escape to infinity, then max-min equals min-max.

Theorem 0.1 (Convex-concave minimax theorem). *Let X be a locally convex topological vector space and let Λ be a convex subset of another locally convex topological vector space. Let*

$$L : X \times \Lambda \rightarrow \mathbb{R} \cup \{\infty, -\infty\}$$

satisfy:

1. for each fixed $\lambda \in \Lambda$, the map $x \mapsto L(x, \lambda)$ is convex and lower semicontinuous;
2. for each fixed $x \in X$, the map $\lambda \mapsto L(x, \lambda)$ is concave;
3. there exists (x_0, λ_0) such that $L(x_0, \lambda_0)$ is finite;
4. one of the usual compactness/interiority conditions holds, for example: there exists $\lambda_0 \in \Lambda$ such that the sublevel sets

$$\{x \in X : L(x, \lambda_0) \leq c\}$$

are compact for all sufficiently large c .

Then

$$\sup_{\lambda \in \Lambda} \inf_{x \in X} L(x, \lambda) = \inf_{x \in X} \sup_{\lambda \in \Lambda} L(x, \lambda).$$

Theorem 0.2 (Minimax for Lagrangians). *For the constrained convex problem*

$$\inf_x f(x) \quad \text{subject to} \quad g_i(x) \leq 0, \quad i = 1, \dots, m,$$

with f and the g_i convex and lower semicontinuous, define

$$L(x, \lambda) = f(x) + \sum_{i=1}^m \lambda_i g_i(x), \quad \lambda \in \mathbb{R}_+^m.$$

If there exists a strictly feasible point x_0 such that

$$g_i(x_0) < 0, \quad i = 1, \dots, m,$$

then

$$\sup_{\lambda \geq 0} \inf_x L(x, \lambda) = \inf_x \sup_{\lambda \geq 0} L(x, \lambda) = \inf \{f(x) : g_i(x) \leq 0, i = 1, \dots, m\}.$$

The strict-feasibility assumption is the usual Slater condition. It is the standard hypothesis in the applications we have been discussing.

Proof. (Sketch using convex-separation.) Let

$$\alpha := \sup_{\lambda \in \Lambda} \inf_{x \in X} L(x, \lambda), \quad \beta := \inf_{x \in X} \sup_{\lambda \in \Lambda} L(x, \lambda).$$

We always know

$$\alpha \leq \beta.$$

Assume for contradiction that

$$\alpha < \beta.$$

Choose a number c strictly between them:

$$\alpha < c < \beta,$$

and define the convex function

$$F(x) := \sup_{\lambda \in \Lambda} L(x, \lambda).$$

Because it is a supremum of convex lower semicontinuous functions of x , F is again convex and lower semicontinuous. The inequality $c < \beta = \inf_x F(x)$ means exactly that the point $(0, c)$ lies strictly below the epigraph of F .

We now have two disjoint convex sets in $X \times \mathbb{R}$:

- the epigraph

$$\text{epi}(F) = \{(x, t) : t \geq F(x)\},$$

- the open half-space

$$X \times (-\infty, c).$$

By the Hahn–Banach separation theorem, there exists a nonzero continuous affine functional on $X \times \mathbb{R}$ separating them. After normalizing, one obtains a continuous linear functional ℓ on X and a scalar a such that

$$\ell(x) + a t \geq d \quad \text{for all } (x, t) \in \text{epi}(F),$$

while

$$\ell(x) + a t < d \quad \text{for all } t < c.$$

Because the second set is vertical in the x -direction, the separator must actually have $a > 0$. Dividing through by a gives an affine minorant of F :

$$F(x) \geq r - \varphi(x)$$

for some continuous affine functional φ and constant r , with strict inequality at the level c .

That affine minorant is the dual certificate. In the abstract convex proof of Fenchel–Moreau, it becomes a subgradient. In the present minimax proof, it is used to manufacture a dual variable λ^* whose lower bound exceeds c :

$$\inf_x L(x, \lambda^*) \geq c.$$

But this contradicts the definition of

$$\alpha = \sup_{\lambda} \inf_x L(x, \lambda) < c.$$

Hence the strict gap is impossible, so $\alpha = \beta$.

In the Lagrangian case, the separating functional acts on the graph or epigraph of the constrained value function. When one unpacks the separator in coordinates adapted to the constraint map

$$x \mapsto (g_1(x), \dots, g_m(x)),$$

its coefficients are exactly the multipliers

$$\lambda = (\lambda_1, \dots, \lambda_m) \in \mathbb{R}_+^m.$$

Thus, the multiplier vector is not inserted by hand at the last moment. It has a geometric source as the normal vector of the separating hyperplane, read in the coordinates of the constraint space. \square