Lecture 22: Lagrangian Duality and the Augmented Lagrangian

Charles A. Bouman

Purdue University

Friday, October 16, 2026

Outline

1. The Lagrangian and the dual function
2. Strong duality: what usually happens
3. The augmented Lagrangian and its algorithm
4. Why the split-variable algorithm works

Reading: Sections 9.2 and 9.3.

The question from last lecture

The split-variable algorithm solves $\arg\min_x \{f(x) + g(x)\}$ by repeating

$$\begin{aligned} (\hat{x}, \hat{v}) &\leftarrow \arg\min_{(x,v)} \left\{ f(x) + g(v) + \tfrac{a}{2}||x - v + u||^2 \right\} \\ u &\leftarrow u + (\hat{x} - \hat{v}) \ . \end{aligned}$$

It was justified by intuition only. Today's questions:

- What is u, and does it converge?
- Does the solution meet the constraint $\hat{x} = \hat{v}$?

The answers come from the theory of constrained optimization, for a general problem that includes the split problem as a special case.

The general constrained problem

The general constrained problem with a linear constraint is

$$x^* = \underset{\substack{x \in \Omega \\ Cx = b}}{\arg\min} f(x) \ ,$$

where

- $f : \mathbb{R}^N \to \mathbb{R}$,
- $\Omega \subset \mathbb{R}^N$ is a closed convex set, and
- $C \in \mathbb{R}^{K \times N}$ and $b \in \mathbb{R}^K$.

Positivity is $\Omega = \mathbb{R}^{+N}$. The split is $C = [\, I \ \ -I \,]$ and $b = 0$.

Also, the minimum may not exist, so we write:

$$p^* = \underset{\substack{x \in \Omega \\ Cx = b}}{\inf} f(x) \ .$$

When the minimum exists, $p^* = f(x^*)$.

The Lagrangian

Add a cost on the constraint error $Cx - b$. Define the **Lagrangian**

$$L(x; \lambda) = f(x) + \lambda^t (Cx - b) \ ,$$

where $\lambda \in \mathbb{R}^K$ is the **Lagrange multiplier**.

For a fixed λ, the constraint is gone:

$$x(\lambda) = \arg\min_{x \in \Omega} L(x; \lambda) \ .$$

The problem is choosing λ. It must satisfy

$$Cx(\lambda) = b \ ,$$

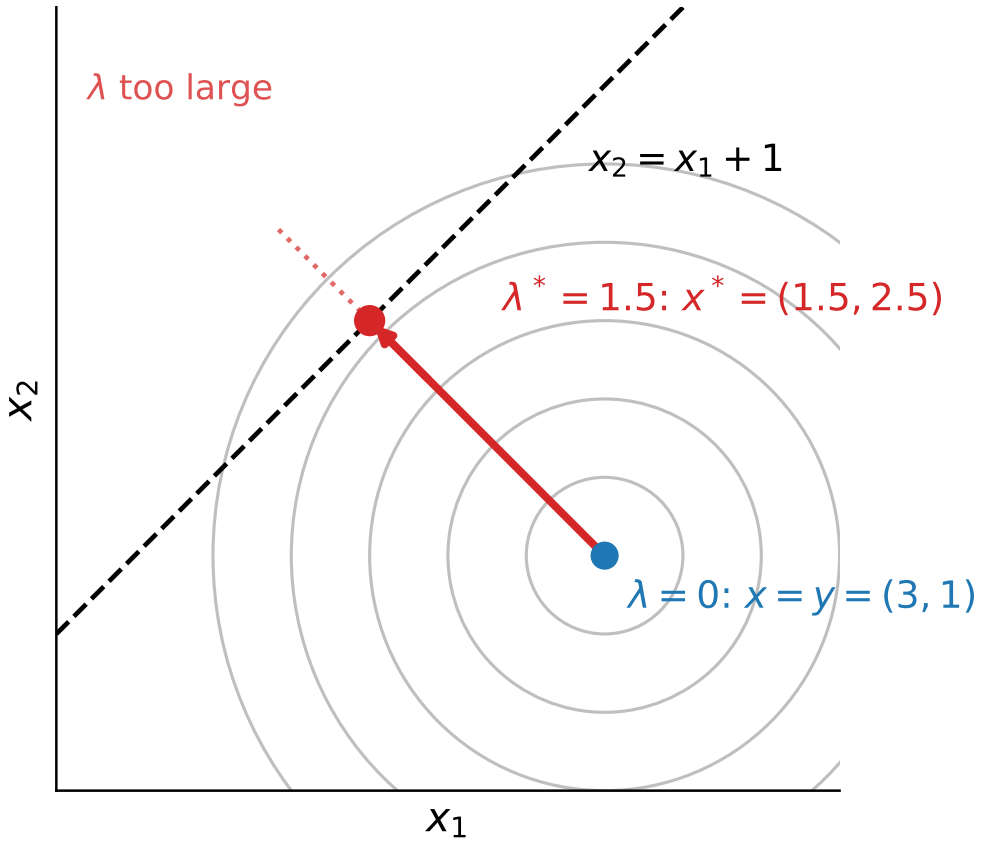
and this equation is usually hard to solve.

What the Lagrange multiplier means

- λ_k is the cost per unit of error in constraint k.
- Raising λ_k pushes the minimizer toward smaller $(Cx - b)_k$.
- Lowering λ_k pushes the minimizer toward larger $(Cx - b)_k$.
- Adjust the cost until the constraint is met.

Example: a line constraint in two dimensions

Minimize $f(x) = \frac{1}{2}\|x - y\|^2$ subject to $x_2 = x_1 + 1$, so $L(x; \lambda) = f(x) + \lambda(x_1 - x_2 + 1)$.



- $x(\lambda) = (y_1 - \lambda,\ y_2 + \lambda)$
- $\lambda = 0$: $x = y$, off the line
- Raising λ slides x to the line
- $\lambda^* = \frac{1}{2}(y_1 - y_2 + 1)$
- $x^* = \left(\frac{y_1+y_2-1}{2},\ \frac{y_1+y_2+1}{2}\right)$
- Too large: x crosses the line

The dual function

Define the **dual function**

$$g(\lambda) = \inf_{x \in \Omega} L(x; \lambda) \ ,$$

where $g : \mathbb{R}^K \to [-\infty, \infty)$.

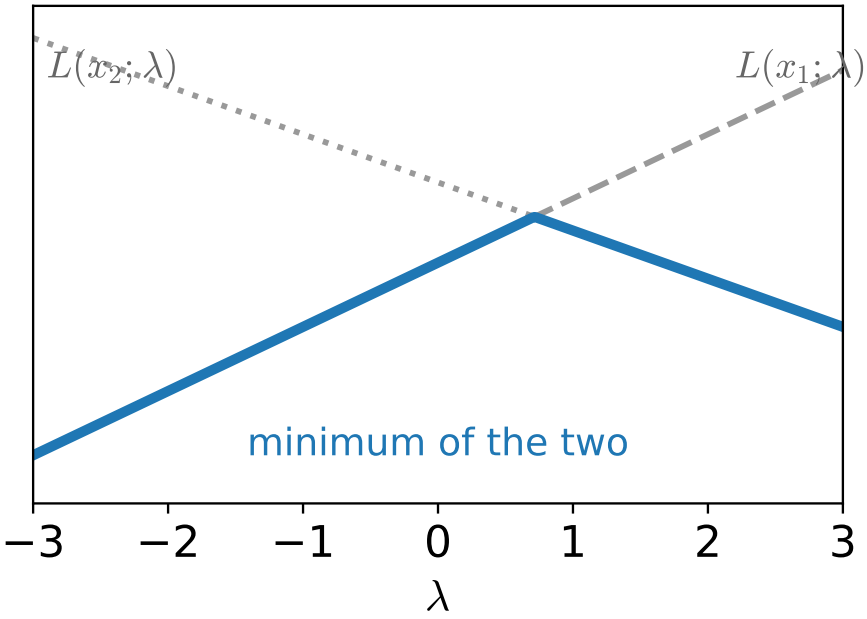
- $f(x)$ is the **primal function**; $g(\lambda)$ is its dual.
- For some λ, $g(\lambda) = -\infty$. This causes no problems.

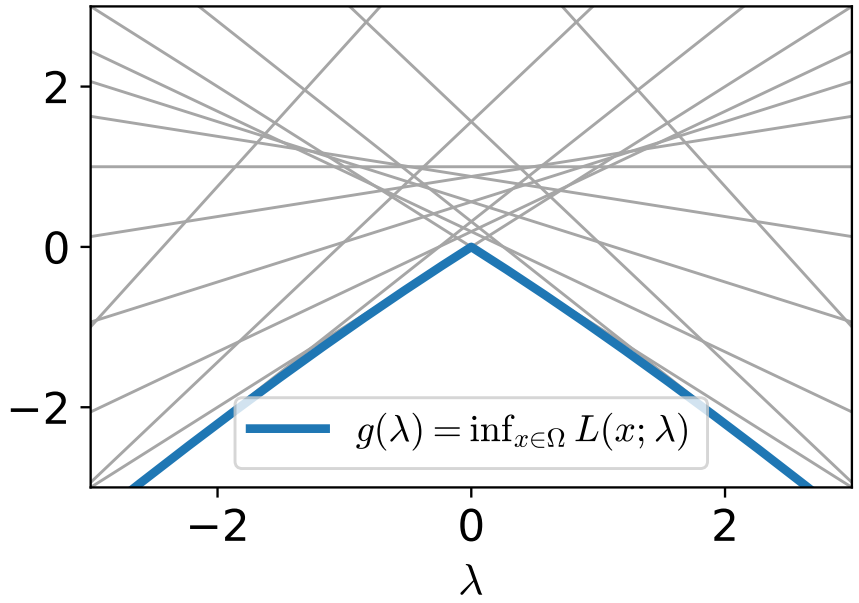
Property: Concavity of the dual function

The dual function $g(\lambda)$ is concave on $\lambda \in \mathbb{R}^K$.

This holds even when $f(x)$ is not convex.

Why the dual function is concave





- For fixed x, $L(x;\lambda)$ is a line in λ, so it is concave.
- The minimum of two concave functions is concave (left).
- The minimum of any number of them is also concave (right).
- $g(\lambda)$ is the minimum over all $x \in \Omega$, so it is concave.

Strong duality

Define d^* as the maximum of the dual function:

$$d^* \triangleq \sup_{\lambda \in \mathbb{R}^K} g(\lambda) \ ,$$

and let λ^* denote a value that achieves it.

Strong duality

$$\inf_{\substack{x \in \Omega \\ Cx = b}} f(x) = \sup_{\lambda \in \mathbb{R}^K} g(\lambda)$$

p^*, the constrained minimum $=$ d^*, the dual maximum

In general $p^* \geq d^*$, but for almost every practical problem they are equal. The next slide says when.

When strong duality holds

Slater's condition: f is convex, and some point strictly inside Ω meets the constraint $Cx = b$. It almost always holds.

Slater's theorem

Under Slater's condition, strong duality holds:

$$p^* = d^* = g(\lambda^*) \quad \text{for some } \lambda^* \in \mathbb{R}^K \ .$$

Property: Optimal Lagrange multiplier

The constrained solution is the unconstrained one with $\lambda = \lambda^*$:

$$x^* = \arg\min_{x \in \Omega} L(x; \lambda^*) \ .$$

The bottom line

The optimal multiplier is $\lambda^* = \arg\max_\lambda g(\lambda)$. But in practice g is usually unknown, so this cannot be done directly.

The augmented Lagrangian

Define the **augmented Lagrangian**

$$L(x; a, \lambda) = f(x) + \frac{a}{2}||Cx - b||^2 + \lambda^t(Cx - b) \ ,$$

where $a > 0$.

- The quadratic term drives the solution toward the constraint.
- It also gives a simple update of λ that converges to λ^*.
- So g is maximized without ever knowing g.

The key fact: one minimization equals one step in the dual

Lemma

Under Slater's condition, for all $\lambda \in \mathbb{R}^K$,

$$\inf_{x \in \Omega} L(x; a, \lambda) = \sup_{\lambda'} \left\{ g(\lambda') - \frac{1}{2a} ||\lambda' - \lambda||^2 \right\} ,$$

and the minimizer x^{**} and maximizer λ^{**} satisfy

$$\lambda^{**} = \lambda + a(Cx^{**} - b) .$$

- Left: one unconstrained minimization over x.
- Right: a step toward the maximum of g, kept near λ.
- The new multiplier is the old one plus a times the constraint error.

The augmented Lagrangian algorithm

Repeat the minimization and the update:

$$\begin{aligned} x^{(k+1)} &= \arg\min_{x\in\Omega} L(x; a, \lambda^{(k)}) \ , \\ \lambda^{(k+1)} &= \lambda^{(k)} + a(Cx^{(k+1)} - b) \ . \end{aligned}$$

Property: Convergence of augmented Lagrangian

Under Slater's condition, $\lambda^{(k)} \to \lambda^*$, and $x^* = \arg\min_{x\in\Omega} L(x; a, \lambda^*)$ solves the constrained problem.

The scaled form: $u = \lambda/a$

Completing the square with $u = \lambda/a$ gives
$L(x; a, \lambda) = f(x) + \frac{a}{2}||Cx - b + u||^2 - \frac{a}{2}||u||^2$, so the algorithm becomes

Augmented Lagrangian Algorithm

Initialize $u \leftarrow \mathbf{0}$

For K iterations $\{$

$$\hat{x} \leftarrow \underset{x \in \Omega}{\arg\min} \left\{ f(x) + \frac{a}{2}||Cx - b + u||^2 \right\}$$

$$u \leftarrow u + C\hat{x} - b$$

$\}$

- u is the Lagrange multiplier scaled by $1/a$, and $u \to u^* = \lambda^*/a$.
- With $C = [\, I \;\; -I \,]$ and $b = 0$, this is the split-variable algorithm.
- So that algorithm converges, and u is the scaled multiplier.

How to choose a

Convergence of the augmented Lagrangian algorithm does not depend on the choice of a.

A large a converges in fewer iterations, for two reasons.

- The penalty $\frac{a}{2}||Cx - b||^2$ pulls x to the constraint faster.
- The dual step's restraint $\frac{1}{2a}||\lambda' - \lambda||^2$ is weaker.

A large a also has a disadvantage.

- The x update becomes more ill conditioned.
- So each iteration is more difficult.

The choice of a balances fast convergence against well-conditioned minimization.

Summary

1. The Lagrangian $L(x;\lambda) = f(x) + \lambda^t(Cx - b)$ puts a cost λ on the constraint error. The right λ makes $Cx(\lambda) = b$.

2. The dual function $g(\lambda) = \inf_{x\in\Omega} L(x;\lambda)$ is concave. Under Slater's condition, $p^* = d^* = g(\lambda^*)$ and $x^* = \arg\min_{x\in\Omega} L(x;\lambda^*)$.

3. With $L(x;a,\lambda) = f(x) + \frac{a}{2}||Cx - b||^2 + \lambda^t(Cx - b)$, one minimization gives the update $\lambda^{(k+1)} = \lambda^{(k)} + a(Cx^{(k+1)} - b)$, and $\lambda^{(k)} \to \lambda^*$ for any a.

4. In scaled form, $u = \lambda/a$: $\hat{x} \leftarrow \arg\min_{x\in\Omega}\{f(x) + \frac{a}{2}||Cx - b + u||^2\}$, $u \leftarrow u + C\hat{x} - b$. With $C = [\,I \;\; -I\,]$ this is the split-variable algorithm.