

Lecture 21: Constrained Optimization; Lagrangian Duality

Charles A. Bouman

Purdue University

Wednesday, October 14, 2026

Outline

- 1 Positivity constraints and the extended function $\tilde{g}$
- 2 Splitting the variable: $x = v$, and the penalty $\frac{a}{2} \|x - v\|^2$
- 3 The correction term u ; the split-variable augmented Lagrangian and ADMM algorithms
- 4 Positivity by clipping
- 5 The general constrained problem and p^*
- 6 The Lagrangian $L(x; \lambda)$ and the dual function $g(\lambda)$
- 7 Duality gap, strong duality, and the optimal Lagrange multiplier
- 8 Slater's conditions and Slater's theorem

Core idea

Constrained optimization is important not only as a way to enforce physical constraints but also as a strategy for making large and difficult optimization problems numerically more tractable. This runs against the natural intuition that unconstrained optimization is easier than constrained optimization.

Today's Big Ideas

- 1 A positivity constraint can be written into an unconstrained cost function by adding a term that is zero inside the constraint set and infinite outside it.
- 2 Splitting one variable into two and requiring the two to be equal, then adding a quadratic penalty and a correction term, gives an unconstrained problem whose solution meets the constraint exactly.
- 3 Splitting is worth doing because the two terms of the cost can then be minimized separately, each by the algorithm best suited to it, which is what ADMM does.
- 4 The Lagrangian converts a constrained problem into an unconstrained one, and its dual function is always concave and never rises above the constrained minimum.
- 5 Strong duality means the largest value of the dual equals the constrained minimum, and Slater's conditions guarantee strong duality for convex problems.

Positivity constraints, and how to write them as a cost

Positivity constraints are very common in imaging problems since the unknown, X , often represents quantities related to **energy or absorption**, which must be positive for physical reasons. The MAP estimate must then be computed with the convex constraint $X \in \mathbb{R}^{+N}$.

Start from the familiar unconstrained problem with two terms,

$$\hat{x} = \arg \min_{x \in \mathbb{R}^N} \{f(x) + g(x)\} ,$$

where $f(x)$ is the negative log likelihood and $g(x)$ is the negative log prior probability.

This same form can enforce positivity by setting $g(x) = \infty$ for $x \notin \mathbb{R}^{+N}$.

The extended function $\tilde{g}$

Choose

$$\tilde{g}(x) = \begin{cases} 0 & \text{if } x \in \mathbb{R}^+{}^N, \\ \infty & \text{if } x \notin \mathbb{R}^+{}^N \end{cases}.$$

Then

$$\arg \min_x \{f(x) + \tilde{g}(x)\} = \arg \min_{x \in \mathbb{R}^+{}^N} f(x).$$

- Adding $\tilde{g}$ enforces positivity by making the cost function infinite when x moves outside the constraint.
- $\tilde{g}$ is an **extended function** because it takes values on $\mathbb{R} \cup \{\infty\}$.
- Perhaps surprisingly, $\tilde{g} : \mathbb{R}^N \rightarrow \mathbb{R} \cup \{\infty\}$ is also a **convex** function if we apply the standard definition of convexity using the convention that $\infty + a = \infty$. It is also proper and closed.

Now something counterintuitive: split the variable

Take a perfectly good unconstrained optimization problem with a single independent variable, x , and “**split**” the variable into two variables, x and v , with the constraint that $x = v$:

$$(\hat{x}, \hat{v}) = \arg \min_{\substack{(x,v) \\ x=v}} \{f(x) + g(v)\} .$$

- Well that certainly seems like a silly thing to do.
- Given the constraint, it must be that $\hat{x} = \hat{v}$, so the solutions to the original and split problems must be the same.
- But why would anyone want to do such a thing?

Rather than answering that now, we solve the constrained problem and see where it leads.

A penalty enforces the constraint, but never exactly

To enforce the constraint, add a term that penalizes large differences between x and v :

$$(\hat{x}, \hat{v}) = \arg \min_{(x,v)} \left\{ f(x) + g(v) + \frac{a}{2} \|x - v\|^2 \right\} ,$$

where a is a constant that controls the gain of the penalty term.

- As a increases, the constraint $x = v$ is more aggressively enforced, so a very large a should force $\hat{x} \approx \hat{v}$.
- However, no matter how large we make a , it will likely never be the case that $\hat{x} = \hat{v}$. So this solution is not exactly the same as that of the original problem.

The correction term u

We need an additional term that can “push” the solution toward a result that exactly meets the constraint $\hat{x} = \hat{v}$:

$$(\hat{x}, \hat{v}) = \arg \min_{(x, v)} \left\{ f(x) + g(v) + \frac{a}{2} \|x - v + u\|^2 \right\} ,$$

where u must be chosen to achieve the desired constraint $\hat{x} = \hat{v}$.

What this expression is

In Section 9.3 we will see that the expression being minimized is known as the **augmented Lagrangian** for our constrained optimization problem, and the term u serves a role equivalent to that of a Lagrange multiplier.

The correct value of u can be determined by a simple algorithm.

Pseudocode: split-variable augmented Lagrangian

Split-Variable Augmented Lagrangian Algorithm

Initialize $u \leftarrow \mathbf{0}$ and $\hat{v} \leftarrow \mathbf{0}$

For K iterations {

$$(\hat{x}, \hat{v}) \leftarrow \arg \min_{(x, v)} \left\{ f(x) + g(v) + \frac{a}{2} \|x - v + u\|^2 \right\}$$

$$u \leftarrow u + (\hat{x} - \hat{v})$$

}

Section 9.2 introduces the machinery of convex optimization, including the duality principle, and Section 9.3 uses that machinery to prove that this algorithm converges to the correct value of u .

Why the update of u drives the constraint

With each iteration, the value of u is increased (or decreased) by the difference $\delta = \hat{x} - \hat{v}$.

- If δ is positive, then $u \leftarrow u + \delta$ increases the value of u , which then “pushes” δ down until $\hat{x} = \hat{v}$.
- If δ is negative, then $u \leftarrow u + \delta$ decreases the value of u , which then “pushes” δ up until $\hat{x} = \hat{v}$.

Either way, this iterative process drives u to a value which causes the constraint $\hat{x} = \hat{v}$ to be met.

Pseudocode: split-variable ADMM

Any engineer worth their salt should have a strong desire to approximately optimize (x, v) in two steps using **alternating minimization** of x and v . It won't produce the same answer, but it should be close, and it seems much easier to do.

Split-Variable ADMM Algorithm

Initialize $u \leftarrow \mathbf{0}$ and $\hat{v} \leftarrow \mathbf{0}$

For K iterations {

$$\hat{x} \leftarrow \arg \min_x \left\{ f(x) + \frac{a}{2} \|x - \hat{v} + u\|^2 \right\}$$

$$\hat{v} \leftarrow \arg \min_v \left\{ g(v) + \frac{a}{2} \|\hat{x} - v + u\|^2 \right\}$$

$$u \leftarrow u + (\hat{x} - \hat{v})$$

}

This is the **alternating directions method of multipliers**, or ADMM. It also converges for convex optimization problems.

Why split the variable at all?

Why solve the original problem this way rather than by direct optimization?

- For many important optimization problems, the two functions, $f(x)$ and $g(x)$, are **so different that a single optimization approach is not appropriate for both**.
- Since the split-variable ADMM algorithm allows the two expressions to be minimized separately, each term may be minimized with a different algorithm that is best suited to its structure.

Intuitively, the ADMM approach allows the two variables, x and v , to move along separate paths and then converge to a single coherent solution.

Positivity: the v update has a closed form

Consider again optimization with a positivity constraint. The optimization over v can be computed in closed form:

$$\begin{aligned}\hat{v} &= \arg \min_v \left\{ \tilde{g}(v) + \frac{a}{2} \|\hat{x} - v + u\|^2 \right\} \\ &= \arg \min_{v \in \mathbb{R}^{+N}} \left\{ \frac{a}{2} \|\hat{x} + u - v\|^2 \right\} \\ &= \text{clip}\{\hat{x} + u, \mathbb{R}^{+N}\} \ ,\end{aligned}$$

where $\text{clip}\{x, \mathcal{A}\}$ projects or “clips” any vector x onto the convex set $\mathcal{A}$.

For nonnegative vectors, $\mathcal{A} = \mathbb{R}^{+N}$ and for each component

$$\left[\text{clip}\{x, \mathbb{R}^{+N}\} \right]_i = \text{clip}\{x_i, [0, \infty)\} \ ,$$

so each component is clipped to be nonnegative.

Pseudocode: split-variable ADMM for a positivity constraint

Split-Variable ADMM for Positivity Constraint

Initialize $u \leftarrow \mathbf{0}$ and $\hat{v} \leftarrow \mathbf{0}$

For K iterations {

$$\hat{x} \leftarrow \arg \min_x \left\{ f(x) + \frac{a}{2} \|x - (\hat{v} - u)\|^2 \right\}$$

$$\hat{v} \leftarrow \text{clip}\{\hat{x} + u, \mathbb{R}^{+N}\}$$

$$u \leftarrow u + (\hat{x} - \hat{v})$$

}

This provides a simple and effective method of enforcing positivity in any convex optimization problem. As a practical matter, it can also be used in nonconvex optimization problems, but its theoretical convergence properties to a global minimum are no longer guaranteed in that case.

Section 9.2: the general constrained problem

The previous section gave intuitive motivation, but the discussion was primarily based on “hand-waving” arguments. We now develop the theory on a solid footing, for problems of the general form

$$x^* = \arg \min_{\substack{x \in \Omega \\ Cx = b}} f(x) , \quad \text{or} \quad p^* = \inf_{\substack{x \in \Omega \\ Cx = b}} f(x) ,$$

where $b \in \mathbb{R}^K$, $C \in \mathbb{R}^{K \times N}$, $\Omega \subset \mathbb{R}^N$ is assumed to be a closed convex set with nonempty interior, and $f(x)$ takes values in $\mathbb{R}$.

- To enforce positivity, set $\Omega = \mathbb{R}^{+N}$, a closed convex set with a nonempty interior.
- This is a convex optimization problem. It is not the most general form: the constraint could be nonlinear, $C(x) = b$, or $f(x)$ could be nonconvex. Nonetheless it covers a wide range of important practical problems, including **linear programming**.
- The first form implicitly assumes that the global minimum exists and is unique, which may not be the case even when $f(x)$ is convex. The second form uses the infimum, with $p^* \in [-\infty, \infty)$. If the function takes on its minimum, then $p^* = f(x^*)$.

The Lagrangian, and the difficulty of choosing λ

The Lagrange method converts a constrained problem to an unconstrained one by adding a linear term that drives the solution to the constraint.

Define the **Lagrangian**

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

where $\lambda \in \mathbb{R}^K$ is the **Lagrange multiplier**, and then solve the unconstrained problem

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

- The difficulty is selecting the correct value for λ .
- On the simplest level, λ must be chosen so that the constraint $Cx(\lambda) = b$ is met.
- This is typically not a practical approach, since it is often difficult to solve this equation analytically or numerically.

The dual function, and its concavity

Define the **dual function**

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

where $g : \mathbb{R}^K \rightarrow [-\infty, \infty)$. For some values of λ it may be that $g(\lambda) = -\infty$; this won't cause any problems. We call $f(x)$ the **primal function** and $g(\lambda)$ its dual.

For each distinct value of x , the function $L(x; \lambda) = f(x) + \lambda^t(Cx - b)$ is an affine function of λ , which is a concave function of λ . So $g(\lambda)$ is formed from the minimum of a set of concave functions.

Property: Concavity of the dual function

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

This concavity property always holds **even when the primal function, $f(x)$, is not convex.**

The duality gap

The dual function is important because it is a lower bound that can be maximized to find the best value for the Lagrange multiplier. Define d^* as the supremum of the dual function:

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

Let λ^* denote a value of the Lagrange multiplier that achieves that maximum. So when λ^* exists, then $d^* = g(\lambda^*)$.

The **duality gap** is $p^* - d^*$: the difference between the minimum of the primal cost function and the maximum of its dual.

Property: Nonnegative duality gap

The duality gap is nonnegative so that $p^* \geq d^*$.

Proof: the duality gap is nonnegative

The duality gap is a direct result of the very general **max-min inequality**.
Applied to the Lagrangian,

$$\inf_{x \in \Omega} \sup_{\lambda \in \mathbb{R}^K} L(x; \lambda) \geq \sup_{\lambda \in \mathbb{R}^K} \inf_{x \in \Omega} L(x; \lambda) .$$

To understand the left-hand side, notice that

$$\sup_{\lambda \in \mathbb{R}^K} L(x; \lambda) = \begin{cases} f(x) & \text{if } Cx = b , \\ \infty & \text{if } Cx \neq b \end{cases} , \quad \text{so} \quad \inf_{x \in \Omega} \sup_{\lambda} L(x; \lambda) = p^* .$$

On the right-hand side, $\inf_{x \in \Omega} L(x; \lambda) = g(\lambda)$. Substituting both results,

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

and the proof is complete.

Strong duality and the optimal Lagrange multiplier

In typical cases the dual is not just a lower bound: the solutions to the primal and dual occur at the same place and take the same value. **Strong duality** holds when $p^* = d^*$. It holds for most practical problems.

Property: Optimal Lagrange multiplier

Consider a constrained optimization problem of the infimum form such that

- strong duality holds with $p^* = d^*$,
- $x^* \in \Omega$ achieves the minimum to the constrained optimization problem, and
- $\lambda^* \in \mathbb{R}^K$ achieves the maximum of the dual function.

Then x^* is also a solution to the unconstrained optimization problem

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

Furthermore, if $f(x)$ is strictly convex, then x^* is unique.

Proof: the optimal Lagrange multiplier

If x^* is the solution to the constrained optimization problem, then $Cx^* - b = \mathbf{0}$. Therefore, when strong duality holds,

$$\begin{aligned} L(x^*; \lambda^*) &= f(x^*) + [\lambda^*]^t (Cx^* - b) \\ &= f(x^*) \\ &= p^* = d^* \\ &= g(\lambda^*) \\ &= \inf_{x \in \Omega} L(x; \lambda^*) . \end{aligned}$$

So when strong duality holds, a solution to the constrained minimization problem, x^* , must also be a solution to the unconstrained optimization problem when λ^* is used as the Lagrange multiplier. Moreover, if $f(x)$ is strictly convex, then this solution to the unconstrained problem must be unique.

Intuition: the continuously differentiable case

Suppose both the primal function and its dual are continuously differentiable. In this highly regular case the gradient of the dual is $\nabla g(\lambda^*) = Cx(\lambda^*) - b$. Since $g(\lambda)$ is concave, λ^* is the global maximum if and only if $\nabla g(\lambda^*) = Cx(\lambda^*) - b = \mathbf{0}$.

From this we can show that $d^* \geq p^*$:

$$\begin{aligned} d^* &= g(\lambda^*) \\ &= \min_{x \in \Omega} \{f(x) + [\lambda^*]^t(Cx - b)\} \\ &= f(x(\lambda^*)) + [\lambda^*]^t(Cx(\lambda^*) - b) \\ &= f(x(\lambda^*)) \\ &\geq f(x^*) = p^* . \end{aligned}$$

Together with the nonnegative duality gap property, this ensures $d^* = p^*$, so strong duality holds when $f(x)$ is sufficiently regular on $\mathbb{R}^N$.

Slater's conditions and Slater's theorem

A relatively general set of technical conditions known as **Slater's conditions** ensures strong duality for just about any reasonable convex optimization problem.

Definition: Slater's Conditions

The optimization problem of the infimum form is said to meet Slater's conditions if there exists an x' in the interior of Ω such that $Cx' = b$.

Stated differently, there must be at least one **strictly feasible solution**: it falls in the interior of Ω and it meets the constraint $Cx' = b$. For most practical problems this is easy to verify.

Theorem: Slater's Theorem

Consider a constrained optimization problem of the infimum form such that Slater's conditions hold and $f(x)$ is a convex function on a closed convex set Ω . Then

- strong duality holds with $p^* = d^*$ and
- there exists $\lambda^* \in \mathbb{R}^K$ such that $d^* = g(\lambda^*)$.

The two-step recipe — and the catch

Slater's theorem tells us how to solve any convex optimization problem when Slater's conditions hold.

- 1 Determine λ^* by solving the dual problem:

$$\lambda^* = \arg \max_{\lambda \in \mathbb{R}^K} g(\lambda) .$$

- 2 Use the resulting λ^* to solve the unconstrained problem

$$x^* = \arg \min_{x \in \Omega} \{ f(x) + [\lambda^*]^t (Cx - b) \} ,$$

where x^* is then the solution to the constrained problem.

The catch

The difficulty with this approach is that the dual function, $g(\lambda)$, typically does not have closed form. So it is still not clear how to determine λ^* . Section 9.3, the augmented Lagrangian, provides a practical solution to this problem.

Summary

- Positivity is written into the cost by the extended function $\tilde{g}(x) = 0$ on $\mathbb{R}^{+N}$ and ∞ elsewhere, which is convex, proper, and closed.
- Splitting x into (x, v) with $x = v$, penalizing $\frac{a}{2} \|x - v + u\|^2$, and updating $u \leftarrow u + (\hat{x} - \hat{v})$ gives the split-variable augmented Lagrangian algorithm. Minimizing over x and v alternately gives ADMM.
- Splitting pays off when f and g need different algorithms. For positivity the v update is closed form, $\hat{v} = \text{clip}\{\hat{x} + u, \mathbb{R}^{+N}\}$.
- For $p^* = \inf_{x \in \Omega, Cx=b} f(x)$, the Lagrangian is $L(x; \lambda) = f(x) + \lambda^t(Cx - b)$ and the dual is $g(\lambda) = \inf_{x \in \Omega} L(x; \lambda)$, which is always concave. The max-min inequality gives $p^* \geq d^* = \sup_{\lambda} g(\lambda)$.
- Strong duality is $p^* = d^*$, and then x^* minimizes $L(x; \lambda^*)$ over Ω . Slater's conditions (a strictly feasible point) plus convexity give strong duality and the existence of λ^* .

Reading: Sections 9.1–9.2. Next lecture: Section 9.3, The Augmented Lagrangian.