

Lecture 25: The ADMM Algorithm — Applications and Convergence

Charles A. Bouman

Purdue University

Friday, October 23, 2026

Outline

- 1 Recap: the general ADMM algorithm
- 2 Total variation regularization: the cost $\frac{1}{2}||y - Ax||_{\Lambda}^2 + ||Bx||_1$
- 3 The split $v = Bx$, and the augmented Lagrangian iteration
- 4 The x update is quadratic; the v update is shrinkage
- 5 Pseudocode: ADMM for total variation regularization
- 6 Enforcing convex constraints: the extended function $\tilde{g}$
- 7 Pseudocode: ADMM with a convex constraint
- 8 Convergence: the canonical constrained problem
- 9 Saddle point solutions, and the saddle point property
- 10 Convergence of the ADMM algorithm

Core idea

ADMM pays off. Total variation regularization becomes a quadratic minimization plus a shrinkage step; a convex constraint becomes a clip onto the set; and the convergence of the whole algorithm rests on saddle points and duality.

Today's Big Ideas

- 1 Total variation regularization is solved by splitting the variable, which turns the problem into a quadratic minimization in one variable and a shrinkage step in the other.
- 2 A convex constraint such as positivity is enforced by splitting the variable and clipping onto the set, so the minimization over x stays unconstrained.
- 3 Any solution of the canonical constrained problem is a saddle point of the unaugmented Lagrangian, because strong duality holds.
- 4 That saddle point is what makes the ADMM iterations converge: the constraint is met in the limit, the cost reaches its minimum value, and the scaled dual variable reaches the optimal Lagrange multiplier.

Recap: the general ADMM algorithm

For the problem $\hat{x} = \arg \min_{x \in \Omega} \{f(x) + g(Bx)\}$:

ADMM Algorithm for Minimizing $f(x) + g(Bx)$:

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

For K iterations {

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

$$\hat{v} \leftarrow \arg \min_{v \in \mathbb{R}^K} \left\{ g(v) + \frac{a}{2} \|v - (B\hat{x} + u)\|^2 \right\}$$

$$u \leftarrow u + B\hat{x} - \hat{v}$$

}

- The first two updates alternately optimize x and v ; the third updates u to enforce the constraint $Bx = v$.
- One subtle but important point: the optimization with respect to v no longer requires enforcement of the constraint $x \in \Omega$.

Total variation regularization: the problem

The MAP cost function for total variation regularized inversion is

$$f(x; y) = \frac{1}{2} \|y - Ax\|_A^2 + \|Bx\|_1 ,$$

where A is an $M \times N$ forward matrix, B is a general $K \times N$ matrix, and $\|z\|_1 = \sum_i |z_i|$ is the L_1 norm of z .

Why this cost has resisted our earlier methods

- As we learned in the ICD convergence theorem, the convergence of ICD optimization depends critically on the MAP cost function being continuously differentiable.
- In fact, most gradient-based optimization techniques depend on continuity of the derivative in order to achieve convergence.
- So computing the MAP estimate in this case seems to be problematic.

The split $v = Bx$, and the augmented Lagrangian

We convert the unconstrained optimization over the single vector x into an optimization over two vectors, $x \in \mathbb{R}^N$ and $v \in \mathbb{R}^K$, with the constraint that $v = Bx$.

So the problem becomes

$$(\hat{x}, \hat{v}) = \arg \min_{\substack{(x,v) \\ Bx=v}} \left\{ \frac{1}{2} \|y - Ax\|_\Lambda^2 + \|v\|_1 \right\}.$$

From this, we can formulate the associated augmented Lagrangian to yield the following iteration:

Repeat{

$$(\hat{x}, \hat{v}) \leftarrow \arg \min_{(x,v)} \left\{ \frac{1}{2} \|y - Ax\|_\Lambda^2 + \|v\|_1 + \frac{a}{2} \|Bx - v + u\|^2 \right\}$$

$$u \leftarrow u + B\hat{x} - \hat{v}$$

}

The x update is quadratic; the v update is a proximal map

To implement ADMM, we must solve the joint optimization above using alternating minimization of x and v .

The function is quadratic in x , so minimization with respect to this variable is straightforward.

Minimization with respect to v yields the update formula

$$v \leftarrow \arg \min_{v \in \mathbb{R}^K} \left\{ \frac{1}{2} \|v - (Bx + u)\|^2 + \frac{1}{a} \|v\|_1 \right\} .$$

- So here we see that optimization with respect to v is **exactly a proximal mapping**, with precisely the same form as the L_1 proximal map of Lecture 23: $\arg \min_x \left\{ \frac{1}{2} \|y - x\|^2 + \lambda \|x\|_1 \right\} = S_\lambda(y)$.

Therefore, we can express the solution in terms of the shrinkage function:

$$v \leftarrow S_{1/a}(Bx + u) .$$

The resulting alternating minimization

Putting this together with the quadratic update of x , we know that the following repetition of alternating minimizations should converge to the global minimum of the augmented Lagrangian optimization:

Repeat{

$$\hat{x} \leftarrow \arg \min_{x \in \mathbb{R}^N} \left\{ \frac{1}{2} \|y - Ax\|_{\Lambda}^2 + \frac{a}{2} \|Bx - \hat{v} + u\|^2 \right\}$$

$$\hat{v} \leftarrow S_{1/a}(B\hat{x} + u)$$

}

- Both updates are things we already know how to compute: a quadratic minimization, and a componentwise shrinkage.

Pseudocode: ADMM for total variation regularization

A single iteration of the alternating minimization gives the ADMM algorithm:

ADMM Algorithm for Total Variation Regularization:

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

For K iterations {

$$\hat{x} \leftarrow \arg \min_{x \in \mathbb{R}^N} \left\{ \frac{1}{2} \|y - Ax\|_\lambda^2 + \frac{a}{2} \|Bx - \hat{v} + u\|^2 \right\}$$

$$\hat{v} \leftarrow S_{1/a}(B\hat{x} + u)$$

$$u \leftarrow u + B\hat{x} - \hat{v}$$

}

- Again, this algorithm is guaranteed to converge, based on the theory in the last part of this lecture.

The payoff

Two gaps left open earlier in the course

- *Lecture 20*: the symmetric bound surrogate requires a potential whose second derivative at zero is bounded. Total variation does not have one, so the surrogate methods of Chapter 8 do not apply.
- *Lecture 23*: for the general cost $\frac{1}{2}\|y - Ax\|_\Lambda^2 + \|Bx\|_1$, the ICD updates do not in general converge to the global minimum. This means that ICD cannot be used to guarantee a solution to important problems such as total variation regularization.

Both gaps close here

ADMM solves this exact cost function with a quadratic minimization and a shrinkage step, and its convergence is guaranteed.

Enforcing convex constraints using ADMM

The ADMM algorithm is particularly useful for enforcing convex constraints such as positivity. Consider the general problem

$$\hat{x} = \arg \min_{x \in \mathcal{A}} f(x) ,$$

where $\mathcal{A} \subset \mathbb{R}^N$ is a nonempty convex set and f is a convex function.

Following the logic of Section 9.1, we can express this problem in the form

$$\hat{x} = \arg \min_x \{f(x) + \tilde{g}(x)\} , \quad \text{where} \quad \tilde{g}(x) = \begin{cases} 0 & \text{if } x \in \mathcal{A} , \\ \infty & \text{if } x \notin \mathcal{A} \end{cases}$$

is a proper, closed, convex function on $\mathbb{R}^N$.

The split $x = v$, and its augmented Lagrangian

In order to enforce the constraint, we first split the variable x into two variables, x and v , such that $x = v$. Formally, we represent the optimization problem as

$$(\hat{x}, \hat{v}) = \underset{\substack{(x,v) \in \mathbb{R}^N \times \mathcal{A} \\ x=v}}{\arg \min} \{f(x) + \tilde{g}(v)\} .$$

From this, we can derive the associated augmented Lagrangian optimization algorithm:

Repeat{

$$(\hat{x}, \hat{v}) \leftarrow \underset{(x,v) \in \mathbb{R}^N \times \mathcal{A}}{\arg \min} \left\{ f(x) + \tilde{g}(v) + \frac{a}{2} \|x - v + u\|^2 \right\}$$

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

}

Replacing the joint optimization with alternating minimization of x and v gives the ADMM algorithm for this problem.

ADMM Algorithm for Minimization with Convex Constraint:

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

For K iterations {

$$\hat{x} \leftarrow \arg \min_{x \in \mathbb{R}^N} \left\{ f(x) + \frac{a}{2} \|x - \hat{v} + u\|^2 \right\}$$

$$\hat{v} \leftarrow \text{clip}\{\hat{x} + u, \mathcal{A}\}$$

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

}

- The first two equations alternately optimize x and v ; the last updates u to enforce the constraint $v = x$.
- Importantly, the $\hat{x}$ update requires only **unconstrained** optimization of x .
- $\text{clip}\{x, \mathcal{A}\}$ projects the vector x onto the convex set $\mathcal{A}$ (Section 9.1).

Convergence: the canonical constrained problem

We finish the chapter with formal results guaranteeing convergence of the ADMM algorithm. The results are based on Boyd et al., with some additional context that adapts them for imaging applications.

Define the canonical constrained optimization

$$p^* = \inf_{\substack{x \in \Omega, z \in \Phi \\ Ax + Bz = b}} \{f(x) + h(z)\} ,$$

where $A \in \mathbb{R}^{K \times N}$, $B \in \mathbb{R}^{K \times L}$, and $b \in \mathbb{R}^K$, with the following three properties:

- $f : \Omega \rightarrow \mathbb{R}$ and $h : \Phi \rightarrow \mathbb{R}$ are convex functions on the closed convex sets $\Omega \subset \mathbb{R}^N$ and $\Phi \subset \mathbb{R}^L$.
- Slater's conditions hold.
- There exists a solution $x^* \in \Omega$ and $z^* \in \Phi$ such that $p^* = f(x^*) + h(z^*)$ and $Ax^* + Bz^* = b$.

Lagrangians and the dual function

Again, we define the augmented Lagrangian as

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

the unaugmented Lagrangian as $L(x, z; \lambda) = L(x, z; 0, \lambda)$, and the associated dual function as

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

Furthermore, the supremum of the dual function is denoted by

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

- Our goal is to find conditions that guarantee convergence of the ADMM algorithm to the global minimum of the constrained optimization problem.

Pseudocode: ADMM for the canonical problem

ADMM Algorithm for Canonical Constrained Optimization:

Initialize $a > 0$, $u \leftarrow \mathbf{0}$, and $\hat{z} \leftarrow \mathbf{0}$

For K iterations {

$$\hat{x} \leftarrow \arg \min_{x \in \Omega} \left\{ f(x) + \frac{a}{2} \|Ax + B\hat{z} - b - u\|^2 \right\}$$

$$\hat{z} \leftarrow \arg \min_{z \in \Phi} \left\{ h(z) + \frac{a}{2} \|A\hat{x} + Bz - b - u\|^2 \right\}$$

$$u \leftarrow u + (A\hat{x} - B\hat{z} - b)$$

}

- This is the same three-update structure as before, now stated for the two-block constraint $Ax + Bz = b$.

Definition: saddle point solution

Saddle Point Solution

A **saddle point solution** is a triple (x^*, z^*, λ^*) such that for all $x \in \Omega$, $z \in \Phi$, and $\lambda \in \mathbb{R}^K$,

$$L(x, z; \lambda^*) \geq L(x^*, z^*; \lambda^*) \geq L(x^*, z^*; \lambda) .$$

Saddle point property

Consider the canonical constrained optimization problem above. Then

- there exists $\lambda^* \in \mathbb{R}^K$ such that $d^* = g(\lambda^*)$, and
- the triple (x^*, z^*, λ^*) is a saddle point solution.

In words: any solution to the canonical optimization problem must be a saddle point solution.

Proof of the saddle point property

- The canonical problem assumes that the problem is convex in both $f(x)$ and $h(z)$ and that Slater's conditions hold. Therefore, by the strong duality theorem of Section 9.2, we know that strong duality holds and there exists a $\lambda^* \in \mathbb{R}^K$ such that $d^* = g(\lambda^*)$.
- Since strong duality holds, then by the optimal Lagrange multiplier property, we know that for all $x \in \Omega$ and $z \in \Phi$, $L(x, z; \lambda^*) \geq L(x^*, z^*; \lambda^*)$, where x^* and z^* are the assumed solutions to the constrained optimization problem.
- Furthermore, since x^* and z^* are the assumed solution, we know that $Ax^* + Bz^* = b$. Therefore, for all $\lambda \in \mathbb{R}^K$, $L(x^*, z^*; \lambda^*) = L(x^*, z^*; \lambda)$.

With these two results, we have that

$$L(x, z; \lambda^*) \geq L(x^*, z^*; \lambda^*) \geq L(x^*, z^*; \lambda) ,$$

and the property is proved. □

Property: convergence of the ADMM algorithm

Convergence of ADMM algorithm

Consider a constrained convex optimization problem with the canonical form above. Then the ADMM iterations converge for this problem in the following precise sense:

$$\begin{aligned}\lim_{k \rightarrow \infty} \left\{ A\hat{x}^{(k)} + B\hat{z}^{(k)} - b \right\} &= \mathbf{0} , \\ \lim_{k \rightarrow \infty} \left\{ f(\hat{x}^{(k)}) + h(\hat{z}^{(k)}) \right\} &= p^* , \\ \lim_{k \rightarrow \infty} u^{(k)} &= \frac{\lambda^*}{a} ,\end{aligned}$$

where $\hat{x}^{(k)}$, $\hat{z}^{(k)}$, and $u^{(k)}$ denote the result of the k^{th} ADMM update.

- *Proof:* x^* and z^* are assumed to exist, so by the saddle point property there exists λ^* making (x^*, z^*, λ^*) a saddle point. Boyd et al. show the stated convergence properties hold when $f(x)$ and $h(x)$ are closed proper convex functions and (x^*, z^*, λ^*) is a saddle point solution.

Summary

- **Total variation:** split $v = Bx$ in $\frac{1}{2}\|y - Ax\|_{\Lambda}^2 + \|Bx\|_1$. The x update is quadratic; the v update is the L_1 proximal map, so $\hat{v} \leftarrow S_{1/a}(B\hat{x} + u)$; then $u \leftarrow u + B\hat{x} - \hat{v}$.
- **Convex constraints:** write the constraint as the extended function $\tilde{g}$, split $x = v$. The x update is unconstrained; the v update is $\hat{v} \leftarrow \text{clip}\{\hat{x} + u, \mathcal{A}\}$.
- **Convergence:** for the canonical problem $p^* = \inf\{f(x) + h(z)\}$ subject to $Ax + Bz = b$, with convex f and h , Slater's conditions, and an existing solution: any solution gives a saddle point (x^*, z^*, λ^*) of the unaugmented Lagrangian.
- Given that saddle point, the ADMM iterates satisfy $Ax^{(k)} + Bz^{(k)} - b \rightarrow \mathbf{0}$, $f + h \rightarrow p^*$, and $u^{(k)} \rightarrow \lambda^*/a$.

Reading: Chapter 9.5. This completes Chapter 9. Next lecture: Chapter 10.1–10.2, plug-and-play.