Lecture 24: Variable Splitting and the ADMM Algorithm

Charles A. Bouman

Purdue University

Wednesday, October 21, 2026

Outline

1. Variable splitting, the augmented Lagrangian, and ADMM
2. Total variation: the v update is shrinkage
3. Convex constraints: the v update is a clip (positivity example)
4. Convergence of ADMM

Core idea

To minimize $f(x) + g(Bx)$, split the variable with the constraint $Bx = v$, apply the augmented Lagrangian, and do a single pass of alternating minimization for each update of the multiplier. The result is the ADMM algorithm.

Reading: Sections 9.1 and 9.5 to 9.5.3, *Foundations of Computational Imaging*.

Today's Big Ideas

- Splitting gives each cost term its own variable, with the same solution.
- Augmented Lagrangian forces the split; alternation separates terms.
- One alternating pass per multiplier update gives ADMM.
- Total variation: a quadratic x update and a shrinkage v update.
- Convex constraints: a free x update, then v clips onto the set.
- For convex problems: ADMM converges, constraint met in the limit.

The general problem

Consider the problem

$$\hat{x} = \arg\min_{x \in \Omega} \{f(x) + g(Bx)\} \ , \tag{9.39}$$

where B is a $K \times N$ matrix and $f(x)$ and $g(v)$ are convex functions.

- f is a function of x, and g is a function of the transformed image Bx.
- Example: f the data term, $g(v) = ||v||_1$: total variation.
- For that cost, ICD does not in general reach the global minimum.

Split the variable

Variable splitting replaces Bx with a new variable v and adds the constraint $Bx = v$:

$$(\hat{x}, \hat{v}) = \underset{\substack{(x,v) \\ Bx=v}}{\arg\min} \{f(x) + g(v)\} \ . \tag{9.40}$$

Why the solutions agree

The constraint forces $\hat{v} = B\hat{x}$. So the solutions of (9.39) and (9.40) are the same.

- The two terms of the cost are now functions of two separate variables.

The augmented Lagrangian of the split problem

The augmented Lagrangian converts (9.40) into the unconstrained problem

$$(\hat{x}, \hat{v}) = \underset{(x,v)}{\arg\min} \left\{ f(x) + g(v) + \frac{a}{2} ||Bx - v + u||^2 \right\} \quad , \tag{9.41}$$

where $u = \lambda/a$ is the scaled Lagrange multiplier.

The augmented Lagrangian algorithm finds the u that gives $B\hat{x} = \hat{v}$:

Repeat {

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

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

}

Alternating minimization over x and v

The joint minimization over (x, v) is done by alternating minimization:

$$\begin{aligned}
&\text{Repeat } \{ \\
&\quad \hat{x} \leftarrow \arg\min_x \left\{ f(x) + \frac{a}{2}||Bx - \hat{v} + u||^2 \right\} \\
&\quad \hat{v} \leftarrow \arg\min_v \left\{ g(v) + \frac{a}{2}||B\hat{x} - v + u||^2 \right\} \\
&\quad \}
\end{aligned}$$

- $g(v)$ drops out of the x update, since it has no x.
- $f(x)$ drops out of the v update, since it has no v.
- Alternation is coordinate descent: it should reach the joint minimum.

One inner iteration is enough

- Running alternation to convergence is an inner loop per u update.
- This nested iteration is possible, but it is computationally expensive.

The single inner iteration

Do one x update and one v update for each update of u.

- Define the **ADMM algorithm**, the alternating direction method of multipliers, as the algorithm that results from a single iteration of alternating minimization.
- ADMM still has guaranteed convergence for convex problems.

The ADMM algorithm

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

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

For K iterations $\{$

$$\hat{x} \leftarrow \arg\min_{x \in \Omega} \left\{ f(x) + \frac{a}{2} ||Bx - (\hat{v} - u)||^2 \right\} \tag{9.46}$$

$$\hat{v} \leftarrow \arg\min_{v \in \mathbb{R}^K} \left\{ g(v) + \frac{a}{2} ||v - (B\hat{x} + u)||^2 \right\} \tag{9.47}$$

$$u \leftarrow u + B\hat{x} - \hat{v} \tag{9.48}$$

$\}$

The roles of the three updates

- The x update (9.46) minimizes the augmented Lagrangian over x.
- The v update (9.47) minimizes the augmented Lagrangian over v.
- The u update (9.48) adjusts the multiplier to enforce $Bx = v$.

Two properties of the v update

- The v update ignores $x \in \Omega$; it minimizes over $\mathbb{R}^K$.
- The v update is the proximal map of g/a, evaluated at $B\hat{x} + u$.

Why ADMM is useful

ADMM results from three steps:

1. Split the variable to separate the two terms of the cost.
2. Apply the augmented Lagrangian to enforce the split.
3. Use a single iteration of alternating minimization.

- Different optimization methods can be used on f and on g.
- The constraint $Bx = v$ is relaxed during the iterations and enforced at convergence. This flexibility can allow the algorithm to converge more efficiently.

Total variation: the problem

The MAP cost for total variation regularized inversion is

$$f(x;y) = \frac{1}{2}||y - Ax||_{\Lambda}^2 + ||Bx||_1 \ , \tag{9.52}$$

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

- The cost is not continuously differentiable.
- ICD convergence depends on a continuously differentiable cost.
- Most gradient methods also need a continuous derivative.

Total variation: split the variable

Split with $v = Bx$, where $x \in \mathbb{R}^N$ and $v \in \mathbb{R}^K$:

$$(\hat{x}, \hat{v}) = \underset{\substack{(x,v) \\ Bx=v}}{\arg\min} \left\{ \frac{1}{2}||y - Ax||_\Lambda^2 + ||v||_1 \right\} \quad . \tag{9.53}$$

The augmented Lagrangian algorithm is then

Repeat $\{$

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

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

$\}$

Total variation: the two updates

The x update

The function is quadratic in x, so the x update is a quadratic minimization.

The v update

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

This is the L_1 proximal map with $\lambda = 1/a$, so its solution is the shrinkage function:

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

Here $S_\lambda(y_i) = \text{sign}(y_i) \max\{|y_i| - \lambda, 0\}$ is applied to each component.

Total variation: the resulting alternating minimization

Together, the two updates give an alternating minimization that should converge to the global minimum of the joint problem:

$$
\begin{aligned}
&\text{Repeat } \{ \\
&\qquad \hat{x} \leftarrow \underset{x \in \mathbb{R}^N}{\arg\min} \left\{ \frac{1}{2} ||y - Ax||_\Lambda^2 + \frac{a}{2} ||Bx - \hat{v} + u||^2 \right\} \\
&\qquad \hat{v} \leftarrow S_{1/a}(B\hat{x} + u) \\
&\qquad \}
\end{aligned}
$$

- Each update is simple: a quadratic minimization, then shrinkage.
- ADMM uses one pass of these two updates for each update of u.

Total variation: 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 $\{$

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

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

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

$\}$

- This algorithm computes the MAP estimate for the total variation cost (9.52), and it is guaranteed to converge.

Convex constraints: the problem

Consider the constrained problem

$$\hat{x} = \arg\min_{x \in \mathcal{A}} f(x) \ , \tag{9.61}$$

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

Write the constraint as the extended function

$$\tilde{g}(x) = \begin{cases} 0 & \text{if } x \in \mathcal{A} \ , \\ \infty & \text{if } x \notin \mathcal{A} \end{cases} \ .$$

Then $\tilde{g}$ is a proper, closed, convex function on $\mathbb{R}^N$, and

$$\hat{x} = \arg\min_{x} \left\{ f(x) + \tilde{g}(x) \right\} \ . \tag{9.62}$$

Convex constraints: the v update is a clip

Split with $x = v$. This is the general problem with $g = \tilde{g}$ and $B = I$.

The v update is then

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

Define the **clip** of a vector x onto the convex set $\mathcal{A}$ as

$$\text{clip}\left\{x, \mathcal{A}\right\} = \arg\min_{y \in \mathcal{A}} ||x - y|| \ .$$

It is the point of $\mathcal{A}$ closest to x.

Convex constraints: the ADMM algorithm

ADMM Algorithm for Minimization with Convex Constraint:

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

For K iterations $\{$

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

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

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

$\}$

- The x update is an **unconstrained** minimization.
- Only the clip enforces the constraint.

Example: the positivity constraint

For positivity, $\mathcal{A} = \mathbb{R}^{+N}$, and the clip acts on each component:

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

Split-Variable ADMM for Positivity Constraint:

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

For K iterations $\{$

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

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

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

$\}$

It can also be used for nonconvex f, but then convergence to the global minimum is not guaranteed.

Convergence of ADMM

Property 9.6: Convergence of ADMM

Let f and g be proper closed convex functions, let Slater's conditions hold, and let a solution (x^*, v^*) of (9.40) exist. Let $p^* = f(x^*) + g(v^*)$, and let λ^* be the optimal Lagrange multiplier. Then the ADMM iterates satisfy

$$\lim_{k\to\infty} \left\{ B\hat{x}^{(k)} - \hat{v}^{(k)} \right\} = \mathbf{0} \ ,$$

$$\lim_{k\to\infty} \left\{ f(\hat{x}^{(k)}) + g(\hat{v}^{(k)}) \right\} = p^* \ ,$$

$$\lim_{k\to\infty} u^{(k)} = \frac{\lambda^*}{a} \ ,$$

where $\hat{x}^{(k)}$, $\hat{v}^{(k)}$, and $u^{(k)}$ are the results of the k^{th} ADMM iteration.

- The book states this property for a more general constrained problem.
- Proof for $x = v$: Plug-and-Play lectures, fixed-point theorem.

Summary

- Splitting $\min_x\{f(x) + g(Bx)\}$ gives $\min_{(x,v)}\{f(x) + g(v)\}$ subject to $Bx = v$, with the same solution.
- The augmented Lagrangian minimizes $f(x) + g(v) + \frac{a}{2}||Bx - v + u||^2$, and alternating minimization updates x and v separately.
- ADMM does one pass per multiplier update:
  $\hat{x} \leftarrow \arg\min_{x\in\Omega}\{f(x) + \frac{a}{2}||Bx - (\hat{v} - u)||^2\}$,
  $\hat{v} \leftarrow \arg\min_v\{g(v) + \frac{a}{2}||v - (B\hat{x} + u)||^2\}$, $u \leftarrow u + B\hat{x} - \hat{v}$.
- For total variation, the x update is quadratic and $\hat{v} \leftarrow S_{1/a}(B\hat{x} + u)$.
- For a convex constraint, the x update is unconstrained and $\hat{v} \leftarrow \text{clip}\{\hat{x} + u, \mathcal{A}\}$.
- For convex problems, $B\hat{x}^{(k)} - \hat{v}^{(k)} \to \mathbf{0}$, $f(\hat{x}^{(k)}) + g(\hat{v}^{(k)}) \to p^*$, and $u^{(k)} \to \lambda^*/a$.