

Lecture 24: Variable Splitting

Charles A. Bouman

Purdue University

Wednesday, October 21, 2026

Outline

- 1 The general problem: minimize $f(x) + g(Bx)$
- 2 Variable splitting: $f(x) + g(v)$ subject to $Bx = v$
- 3 The augmented Lagrangian form of the split problem
- 4 Alternating minimization of the joint (x, v) problem
- 5 The nesting problem, and the single-iteration shortcut
- 6 The general ADMM algorithm
- 7 The three tricks, and the practical advantages

Core idea

To minimize $f(x) + g(Bx)$: **split** into $f(x) + g(v)$ with the constraint $Bx = v$, apply the **augmented Lagrangian**, and use a **single iteration** of alternating minimization per outer iteration. That single-iteration shortcut is the ADMM algorithm.

Today's Big Ideas

- 1 Many optimization problems have a cost made of two terms, one a function of the image and the other a function of a linear transformation of the image, and ICD does not always converge to the global minimum of such a cost.
- 2 Variable splitting replaces the transformed image by a second variable and adds the constraint that the two agree, which puts the two terms of the cost in separate variables without changing the solution.
- 3 The augmented Lagrangian turns the split problem back into an unconstrained minimization, and the joint minimization over the two split variables is done by minimizing over one of them at a time.
- 4 One pass of alternating minimization inside each augmented Lagrangian step is enough, and the algorithm that results is the ADMM algorithm, which still has guaranteed convergence.
- 5 The ADMM updates are separate optimizations, so different methods can be used on the two parts of the cost, and only the update of the image is required to satisfy the constraint on the image.

Two synergistic methods

In this section, we will show how constrained optimization can be used to solve a number of practical problems in MAP optimization by using two synergistic methods:

- **variable splitting**, and
 - the **ADMM algorithm** [Fortin and Glowinski, 1983; Gabay, 1983].
-
- In fact, the two methods were informally introduced in Section 9.1 in the context of enforcing positivity constraints.
 - However, in this section, we will re-examine these two methods in a context that is a bit more general.

The general problem

Consider any optimization problem with the general form

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

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

- f is a function of x directly; g is a function of the transformed variable Bx .
- From Lecture 23: with $g(\cdot) = \|\cdot\|_1$ this is exactly the cost $f(x) + \|Bx\|_1$ for which ICD does not converge to the global minimum.

Trick 1: split the variable

Now to solve this problem, we will split the variable x into two variables, x and v , with the constraint that $Bx = v$. So then the new problem is given by

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

Why the two problems are the same

Given the constraint, it must be that $\hat{v} = B\hat{x}$, so the solutions to the two problems must be the same.

- The point of the split: f and g are now functions of two separate variables, so they appear in separate terms of the cost.

Trick 2: apply the augmented Lagrangian

Using the augmented Lagrangian methods of Section 9.3, we can convert this constrained optimization problem into an unconstrained optimization problem with the form

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

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

- From Lecture 22: this is the augmented Lagrangian in the scaled form, with $u = \lambda/a$, and with the constraint $Cx - b = \mathbf{0}$ replaced by $Bx - v = \mathbf{0}$.

The augmented Lagrangian iteration

Now in order to determine the correct value of the Lagrange multiplier u , we will use the augmented Lagrangian algorithm of Section 9.3. This results in the following iteration:

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

- The first line is a **joint** minimization over the pair (x, v) .
- The second line is the usual augmented Lagrangian update of u .

Solving the joint minimization by alternating minimization

Now let us focus on solving the joint optimization over (x, v) . In order to solve this optimization, we can use the method of alternating minimization specified by the following iteration:

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

- Each line minimizes over one of the two split variables while holding the other fixed.

Why each term drops out of the other update

- Notice that we drop the term $g(v)$ in the x update since it is not a function of x .
- And we drop the term $f(x)$ in the v update since it is not a function of v .

Convergence of the alternating iteration

Since this alternating iteration can be a form of coordinate descent from the ICD convergence theorem (Theorem 7.2), we know that it should converge to the desired solution of the augmented Lagrangian minimization.

- This is the payoff of the split: the two terms of the cost are now handled in two separate minimizations.

The nesting problem

Going back to the problem of the augmented Lagrangian, we must embed this alternating minimization within each iteration of the augmented Lagrangian algorithm.

Nested iteration

- Outer loop: the augmented Lagrangian update of u .
 - Inner loop: alternating minimization over x and v , run to convergence.
-
- This would result in a nested iteration, which, while possible, can be computationally expensive.

Trick 3: one inner iteration is enough

Instead, we assume that one iteration of the alternating minimization is enough.

This is not just an approximation

- The **ADMM algorithm** is the algorithm that results from a single iteration of alternating minimization.
- It can be proved that ADMM still has guaranteed convergence [Boyd, Parikh, Chu, Peleato, and Eckstein, 2011; Eckstein, 2012].
- So the inner loop disappears: one x update, one v update, and one u update per iteration.

The general 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 {

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

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

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

}

The first two equations provide alternating optimization of the variables x and v , while the last equation updates u according to the augmented Lagrangian method to enforce the constraint that $Bx = v$.

The roles of the three updates

The algorithm consists of three updates performed repeatedly until convergence.

- The **first update** minimizes the augmented Lagrangian cost function with respect to the variable x .
- The **second update** does the same for v .
- The **third update** updates the variable u of the augmented Lagrangian in order to enforce the required constraint that $v = Bx$.
- Note the form of the v update: $g(v)$ plus $\frac{a}{2} \|v - (B\hat{x} + u)\|^2$.
- Next lecture: for $g = \|\cdot\|_1$, this v update is exactly a proximal mapping of the form solved in Lecture 23.

A subtle but important point

The constraint $x \in \Omega$ appears in only one update

One subtle but important point is that the optimization with respect to v no longer requires the enforcement of the constraint $x \in \Omega$.

- In the ADMM algorithm, the x update minimizes over $x \in \Omega$, while the v update minimizes over all of $\mathbb{R}^K$.
- As we will soon see, this can provide an important simplification in some circumstances.

ADMM results from a series of three tricks

- 1 Split a single variable into two variables in order to separate the two terms in the optimization.
 - 2 Apply the augmented Lagrangian method to enforce consistency of the split variables.
 - 3 Use a single iteration of alternating minimization to minimize with respect to each of the split variables.
-
- Based on this view, it is clear that ADMM can be used for a wide range of applications.

Practical advantages

Intuitively, the ADMM algorithm has many practical advantages.

- **First**, it allows different optimization methods to be used with different parts of the cost function.
- **Second**, the augmented Lagrangian method allows the constraint to be relaxed during the optimization process but ultimately enforces it at convergence.

Intuitively, this flexibility to relax the constraint during optimization can allow the optimization algorithm to converge more efficiently to its solution.

Summary

- General problem: $\hat{x} = \arg \min_{x \in \Omega} \{f(x) + g(Bx)\}$, with B a $K \times N$ matrix and f, g convex.
- Split: $(\hat{x}, \hat{v}) = \arg \min_{(x,v)_{Bx=v}} \{f(x) + g(v)\}$; the constraint forces $\hat{v} = B\hat{x}$, so the solutions are the same.
- Augmented Lagrangian: minimize $f(x) + g(v) + \frac{a}{2} \|Bx - v + u\|^2$ jointly over (x, v) , then update $u \leftarrow u + B\hat{x} - \hat{v}$.
- The joint minimization is done by alternating minimization; $g(v)$ drops out of the x update and $f(x)$ drops out of the v update.
- A single iteration of the alternating minimization gives the ADMM algorithm, which still has guaranteed convergence.
- In the ADMM algorithm, only the x update carries the constraint $x \in \Omega$. The two updates are separate optimizations, so different methods can be used on f and on g .

Reading: Section 9.5. Next lecture: ADMM applied to total variation regularization and to convex constraints, plus the convergence theory.