Statistical Optimization: Lecture 13

Alternating Direction Method of Multipliers

Zijian Guo

Zhejiang University
Center for Data Science

April 21, 2026

From Lecture 12 to Lecture 13

In Lecture 12, we considered the equality-constrained problem

$$\min_{\theta} \; f_0(\theta) \qquad \text{s.t.} \quad h_i(\theta) = 0, \quad i \in \mathcal{E}.$$

In Lecture 13, we do *not* change the basic framework. We only move to a more structured case.

We split the variable into two blocks

$$\theta = (\theta_1, \theta_2),$$

and write the objective and constraint in a more explicit form.

So Lecture 13 should be viewed as a structured special case of the equality-constrained problem from Lecture 12.

Writing h_i explicitly

Suppose the objective is separable:

$$f_0(\theta) = f_1(\theta_1) + f_2(\theta_2).$$

Now write the equality constraint explicitly as

$$h(\theta_1, \theta_2) = A_1\theta_1 + A_2\theta_2 - b = 0.$$

Equivalently, if $h : \mathbb{R}^{n_1+n_2} \to \mathbb{R}^m$, then each equality constraint in Lecture 12 is just one component:

$$h_i(\theta_1, \theta_2) = \left[A_1\theta_1 + A_2\theta_2 - b\right]_i, \qquad i = 1, \ldots, m.$$

So the notation changes only because here the constraint is linear and explicit.

Why this motivates ADMM

In ALM, the primal step is

$$(\theta_1^{k+1}, \theta_2^{k+1}) \approx \arg\min_{\theta_1,\theta_2} L_A(\theta_1, \theta_2, \nu^k; \mu).$$

Even if $f_1(\theta_1) + f_2(\theta_2)$ is separable, the term

$$\frac{\mu}{2} \|A_1\theta_1 + A_2\theta_2 - b\|_2^2$$

still couples the two blocks.

Therefore, the ALM subproblem is usually not separable in θ_1 and θ_2.

This is the motivation for ADMM:

- keep the same augmented Lagrangian as ALM,
- but replace the joint minimization by alternating block updates.

Outline

ADMM: Alternating Direction Method of Multipliers

Convergence proof

Problem form

Consider

$$\begin{aligned} \min_{\theta_1,\theta_2} \quad & f_1(\theta_1) + f_2(\theta_2) \\ \text{s.t.} \quad & A_1\theta_1 + A_2\theta_2 = b. \end{aligned} \tag{1}$$

The objective is separable in θ_1, θ_2, while the two blocks are coupled by the linear constraint.

This is the basic problem form for ADMM.

Why not ALM?

For problem (1), the augmented Lagrangian method solves

$$\min_{\theta_1,\theta_2} L_A(\theta_1, \theta_2, \nu; \mu)$$

jointly in the two blocks.

The difficulty is that the quadratic penalty

$$\frac{\mu}{2}\|A_1\theta_1 + A_2\theta_2 - b\|_2^2$$

couples θ_1 and θ_2, so the subproblem is usually not separable.

ADMM keeps the same augmented Lagrangian, but updates the two blocks *alternately*.

Augmented Lagrangian

For problem (1), define

$$L_A(\theta_1, \theta_2, \nu; \mu) := f_1(\theta_1) + f_2(\theta_2) + \nu^\top (A_1\theta_1 + A_2\theta_2 - b) + \frac{\mu}{2}\|A_1\theta_1 + A_2\theta_2 - b\|_2^2, \qquad \mu > 0. \tag{2}$$

This is exactly the same augmented Lagrangian as in ALM.

The difference is only in the update rule: ALM minimizes L_A jointly in (θ_1, θ_2), while ADMM minimizes it block by block.

ADMM updates

ADMM applies alternating minimization to the augmented Lagrangian:

$$\theta_1^{k+1} := \arg\min_{\theta_1} L_A(\theta_1, \theta_2^k, \nu^k; \mu). \tag{3}$$

$$\theta_2^{k+1} := \arg\min_{\theta_2} L_A(\theta_1^{k+1}, \theta_2, \nu^k; \mu). \tag{4}$$

$$\nu^{k+1} = \nu^k + \mu\left(A_1\theta_1^{k+1} + A_2\theta_2^{k+1} - b\right). \tag{5}$$

So each iteration has three steps: update θ_1, then update θ_2, then update the multiplier.

Where does the ν-update come from?

Define the augmented dual function

$$g_\mu(\nu) := \inf_{\theta_1,\theta_2} L_A(\theta_1, \theta_2, \nu; \mu). \tag{6}$$

If $(\widehat{\theta}_1(\nu), \widehat{\theta}_2(\nu))$ minimizes L_A for a fixed ν, then we calculate gradient (by Danskin's theorem),

$$\nabla g_\mu(\nu) = A_1\widehat{\theta}_1(\nu) + A_2\widehat{\theta}_2(\nu) - b. \tag{7}$$

So a gradient ascent step for maximizing $g_\mu(\nu)$ is

$$\nu^{k+1} = \nu^k + \mu\,\nabla g_\mu(\nu^k) = \nu^k + \mu\big(A_1\widehat{\theta}_1(\nu^k) + A_2\widehat{\theta}_2(\nu^k) - b\big). \tag{8}$$

In the method of multipliers, $(\widehat{\theta}_1, \widehat{\theta}_2)$ is the joint minimizer. In ADMM, we use the current alternating updates $(\theta_1^{k+1}, \theta_2^{k+1})$, which gives

$$\nu^{k+1} = \nu^k + \mu\big(A_1\theta_1^{k+1} + A_2\theta_2^{k+1} - b\big). \tag{9}$$

Why alternating helps

Compare ADMM with ALM for the same problem.

ALM step:

$$(\theta_1^{k+1}, \theta_2^{k+1}) := \arg\min_{\theta_1,\theta_2} L_A(\theta_1, \theta_2, \nu^k; \mu).$$

ADMM steps:

$$\theta_1^{k+1} := \arg\min_{\theta_1} L_A(\theta_1, \theta_2^k, \nu^k; \mu),$$

$$\theta_2^{k+1} := \arg\min_{\theta_2} L_A(\theta_1^{k+1}, \theta_2, \nu^k; \mu).$$

Interpretation. ADMM replaces the joint minimization by a single Gauss–Seidel pass over the two blocks. This is what makes the method much easier to implement in structured problems.

Residual

For the coupling constraint, define the residual

$$r(\theta_1, \theta_2) := A_1\theta_1 + A_2\theta_2 - b. \tag{6}$$

Then the augmented Lagrangian can be written as

$$\mathcal{L}_\mu(\theta_1, \theta_2, \nu) = f_1(\theta_1) + f_2(\theta_2) + \nu^\top r(\theta_1, \theta_2) + \frac{\mu}{2}\|r(\theta_1, \theta_2)\|_2^2. \tag{7}$$

Now define the scaled dual variable

$$u := \frac{1}{\mu}\nu. \tag{8}$$

Scaled form

Using $\nu = \mu u$ and $r = r(\theta_1, \theta_2)$, we complete the square:

$$\nu^\top r + \frac{\mu}{2}\|r\|_2^2 = \frac{\mu}{2}\|r + u\|_2^2 - \frac{\mu}{2}\|u\|_2^2. \tag{9}$$

Therefore,

$$\mathcal{L}_\mu(\theta_1, \theta_2, u) = f_1(\theta_1) + f_2(\theta_2) + \frac{\mu}{2}\|r + u\|_2^2 - \frac{\mu}{2}\|u\|_2^2. \tag{10}$$

For the θ_1- and θ_2-subproblems, the last term is constant, so we may drop it.

Scaled ADMM

The scaled ADMM iteration is

$$\theta_1^{k+1} := \arg\min_{\theta_1} \left\{ f_1(\theta_1) + \frac{\mu}{2} \| A_1\theta_1 + A_2\theta_2^k - b + u^k \|_2^2 \right\}, \tag{11}$$

$$\theta_2^{k+1} := \arg\min_{\theta_2} \left\{ f_2(\theta_2) + \frac{\mu}{2} \| A_1\theta_1^{k+1} + A_2\theta_2 - b + u^k \|_2^2 \right\}, \tag{12}$$

$$u^{k+1} = \nu^k + r^{k+1}, \qquad r^{k+1} := A_1\theta_1^{k+1} + A_2\theta_2^{k+1} - b. \tag{13}$$

Hence

$$u^k = u^0 + \sum_{j=1}^{k} r^j, \tag{14}$$

so u^k is the running sum of the primal residuals.

Theorem (Convergence of ADMM)

Assume that $f_1 : \mathbb{R}^{n_1} \to \mathbb{R}$ and $f_2 : \mathbb{R}^{n_2} \to \mathbb{R}$ are closed, convex, and differentiable, and that the unaugmented Lagrangian has a saddle point. More precisely, there exists $(\theta_1^\star, \theta_2^\star, \nu^\star)$ such that for all θ_1, θ_2, ν,

$$L_0(\theta_1^\star, \theta_2^\star, \nu) \leq L_0(\theta_1^\star, \theta_2^\star, \nu^\star) \leq L_0(\theta_1, \theta_2, \nu^\star).$$

Then the ADMM iterates satisfy

- ***residual convergence:***
$$r^k \to 0;$$
- ***objective convergence:***
$$f_1(\theta_1^k) + f_2(\theta_2^k) \to p^\star;$$
- ***dual variable convergence:***
$$\nu^k \to \nu^\star.$$

Remarks on the theorem

- **ADMM can be slow.** ADMM is often attractive when moderate accuracy is enough. If very high accuracy is required, its convergence may be slow.
- **Intuition for the saddle point assumption.** The primal variables minimize L_0 at fixed $\nu^\star$, while the multiplier $\nu^\star$ maximizes L_0 at fixed $(\theta_1^\star, \theta_2^\star)$. So the Lagrangian looks like a valley in the primal directions and a peak in the dual direction.
- **What is not guaranteed.** The theorem shows residual convergence, objective convergence, and dual variable convergence. But θ_1^k and θ_2^k need not themselves converge to optimal values, unless additional assumptions are imposed.

Optimality conditions

The unaugmented Lagrangian is

$$L_0(\theta_1, \theta_2, \nu) = f_1(\theta_1) + f_2(\theta_2) + \nu^\top(A_1\theta_1 + A_2\theta_2 - b).$$

The necessary and sufficient optimality conditions are:

Primal feasibility

$$A_1\theta_1^\star + A_2\theta_2^\star - b = 0, \tag{15}$$

Dual feasibility

$$\nabla f_1(\theta_1^\star) + A_1^\top \nu^\star = 0, \tag{16}$$

$$\nabla f_2(\theta_2^\star) + A_2^\top \nu^\star = 0. \tag{17}$$

The θ_2-update

By definition, θ_2^{k+1} minimizes

$$L_\mu(\theta_1^{k+1}, \theta_2, \nu^k)$$

with respect to θ_2. Therefore,

$$\nabla f_2(\theta_2^{k+1}) + A_2^\top \nu^k + \mu A_2^\top \big(A_1\theta_1^{k+1} + A_2\theta_2^{k+1} - b\big) = 0. \tag{18}$$

Using the multiplier update

$$\nu^{k+1} = \nu^k + \mu\big(A_1\theta_1^{k+1} + A_2\theta_2^{k+1} - b\big),$$

we obtain

$$\nabla f_2(\theta_2^{k+1}) + A_2^\top \nu^{k+1} = 0. \tag{19}$$

So the second dual feasibility condition is satisfied exactly at every iteration.

The θ_1-update

By definition, θ_1^{k+1} minimizes

$$L_\mu(\theta_1, \theta_2^k, \nu^k)$$

with respect to θ_1. Therefore,

$$\nabla f_1(\theta_1^{k+1}) + A_1^\top \nu^k + \mu A_1^\top \left(A_1\theta_1^{k+1} + A_2\theta_2^k - b\right) = 0. \tag{20}$$

Add and subtract $A_2\theta_2^{k+1}$ inside the last term:

$$\begin{aligned} 0 = \nabla f_1(\theta_1^{k+1}) &+ A_1^\top \nu^k \\ &+ \mu A_1^\top \left(A_1\theta_1^{k+1} + A_2\theta_2^{k+1} - b\right) + \mu A_1^\top A_2(\theta_2^k - \theta_2^{k+1}). \end{aligned} \tag{21}$$

The θ_1-update

Using

$$\nu^{k+1} = \nu^k + \mu\big(A_1\theta_1^{k+1} + A_2\theta_2^{k+1} - b\big),$$

equation (21) becomes

$$\mu A_1^\top A_2(\theta_2^{k+1} - \theta_2^k) = \nabla f_1(\theta_1^{k+1}) + A_1^\top \nu^{k+1}. \tag{22}$$

This motivates the definition of the **dual residual**

$$s^{k+1} := \mu A_1^\top A_2(\theta_2^{k+1} - \theta_2^k). \tag{23}$$

If s^{k+1} is small, then the first dual feasibility condition is nearly satisfied.

Primal and dual residuals

The **primal residual** is the constraint violation

$$r^{k+1} := A_1\theta_1^{k+1} + A_2\theta_2^{k+1} - b. \tag{24}$$

The **dual residual** is

$$s^{k+1} := \mu A_1^\top A_2(\theta_2^{k+1} - \theta_2^{k}). \tag{25}$$

So the optimality conditions are monitored by two quantities:

- r^{k+1}: how well the primal constraint is satisfied;
- s^{k+1}: how well the first dual feasibility condition is satisfied.

Moreover, the second dual feasibility condition

$$\nabla f_2(\theta_2^{k+1}) + A_2^\top \nu^{k+1} = 0$$

already holds exactly at every step.

Stopping criterion

A standard practical stopping test is

$$\|r^k\|_2 \le \varepsilon_{\mathrm{pri}}, \qquad \|s^k\|_2 \le \varepsilon_{\mathrm{dual}}. \tag{26}$$

The tolerances are often chosen by combining an absolute and a relative term:

$$\varepsilon_{\mathrm{pri}} = \varepsilon_{\mathrm{abs}} + \varepsilon_{\mathrm{rel}} \max\{\|A_1\theta_1^k\|_2, \|A_2\theta_2^k\|_2, \|b\|_2\}, \tag{27}$$

$$\varepsilon_{\mathrm{dual}} = \varepsilon_{\mathrm{abs}} + \varepsilon_{\mathrm{rel}}\|A_1^\top \nu^k\|_2. \tag{28}$$

So ADMM stops only when both feasibility and stationarity are sufficiently accurate.

Outline

ADMM: Alternating Direction Method of Multipliers

Convergence proof

Proof setup

Assume the conditions of the convergence theorem hold, and let $(\theta_1^\star, \theta_2^\star, \nu^\star)$ be a saddle point of L_0.
Define

$$p^k := f_1(\theta_1^k) + f_2(\theta_2^k), \qquad r^k := A_1\theta_1^k + A_2\theta_2^k - b,$$

and

$$s^k := \mu A_1^\top A_2(\theta_2^k - \theta_2^{k-1}).$$

We also introduce the Lyapunov function

$$V^k := \frac{1}{\mu}\|\nu^k - \nu^\star\|_2^2 + \mu\|A_2(\theta_2^k - \theta_2^\star)\|_2^2. \tag{A.0}$$

Our goal is to show that V^k decreases along the ADMM iterates.

Three key inequalities

The proof is based on the following three inequalities:

$$V^{k+1} \leq V^k - \mu\|r^{k+1}\|_2^2 - \mu\|A_2(\theta_2^{k+1} - \theta_2^k)\|_2^2, \tag{A.1}$$

$$\begin{aligned} p^{k+1} - p^\star \leq & -(\nu^{k+1})^\top r^{k+1} \\ & - \mu\big(A_2(\theta_2^{k+1} - \theta_2^k)\big)^\top \big(-r^{k+1} + A_2(\theta_2^{k+1} - \theta_2^\star)\big), \end{aligned} \tag{A.2}$$

$$p^\star - p^{k+1} \leq (\nu^\star)^\top r^{k+1}. \tag{A.3}$$

We prove $(A.3)$, then $(A.2)$, and finally $(A.1)$.

Consequences of $(A.1)$

By iterating $(A.1)$, we obtain

$$\mu \sum_{k=0}^{\infty} \left(\|r^{k+1}\|_2^2 + \|A_2(\theta_2^{k+1} - \theta_2^k)\|_2^2 \right) \leq V^0. \tag{A.4}$$

Therefore,

$$r^k \to 0, \qquad A_2(\theta_2^{k+1} - \theta_2^k) \to 0.$$

Since

$$s^{k+1} = \mu A_1^\top A_2(\theta_2^{k+1} - \theta_2^k),$$

it follows that

$$s^k \to 0.$$

So $(A.1)$ already gives primal residual convergence and dual residual convergence.

Objective gap bound

From $(A.2)$, observe that

$$-r^{k+1} + A_2(\theta_2^{k+1} - \theta_2^\star) = -A_1(\theta_1^{k+1} - \theta_1^\star),$$

because

$$r^{k+1} = A_1\theta_1^{k+1} + A_2\theta_2^{k+1} - b$$

and

$$A_1\theta_1^\star + A_2\theta_2^\star - b = 0.$$

Hence $(A.2)$ can be rewritten as

$$p^{k+1} - p^\star \leq -(\nu^{k+1})^\top r^{k+1} + (\theta_1^{k+1} - \theta_1^\star)^\top s^{k+1}. \tag{A.5}$$

This is the bound used to motivate the stopping criterion.

Proof of $(A.3)$

Since $(\theta_1^\star, \theta_2^\star, \nu^\star)$ is a saddle point of L_0, we have

$$L_0(\theta_1^\star, \theta_2^\star, \nu^\star) \le L_0(\theta_1^{k+1}, \theta_2^{k+1}, \nu^\star).$$

Using feasibility of the optimal point,

$$A_1\theta_1^\star + A_2\theta_2^\star - b = 0,$$

the left-hand side is simply $p^\star$. Therefore,

$$p^\star \le f_1(\theta_1^{k+1}) + f_2(\theta_2^{k+1}) + (\nu^\star)^\top r^{k+1}.$$

Since

$$p^{k+1} = f_1(\theta_1^{k+1}) + f_2(\theta_2^{k+1}),$$

we obtain

$$p^\star - p^{k+1} \le (\nu^\star)^\top r^{k+1},$$

which is exactly $(A.3)$.

Proof of $(A.2)$: the θ_1-step

By definition, θ_1^{k+1} minimizes

$$L_\mu(\theta_1, \theta_2^k, \nu^k).$$

So the first-order optimality condition is

$$\nabla f_1(\theta_1^{k+1}) + A_1^\top \nu^k + \mu A_1^\top (A_1\theta_1^{k+1} + A_2\theta_2^k - b) = 0. \tag{A.6}$$

Using

$$\nu^{k+1} = \nu^k + \mu r^{k+1},$$

we rewrite $(A.6)$ as

$$\nabla f_1(\theta_1^{k+1}) + A_1^\top \left(\nu^{k+1} - \mu A_2(\theta_2^{k+1} - \theta_2^k)\right) = 0. \tag{A.7}$$

Therefore, θ_1^{k+1} minimizes

$$f_1(\theta_1) + \left(\nu^{k+1} - \mu A_2(\theta_2^{k+1} - \theta_2^k)\right)^\top A_1\theta_1.$$

Proof of $(A.2)$: the θ_2-step

Similarly, θ_2^{k+1} minimizes

$$L_\mu(\theta_1^{k+1}, \theta_2, \nu^k),$$

so

$$\nabla f_2(\theta_2^{k+1}) + A_2^\top \nu^k + \mu A_2^\top r^{k+1} = \nabla f_2(\theta_2^{k+1}) + A_2^\top \nu^{k+1} = 0. \tag{A.8}$$

Hence θ_2^{k+1} minimizes

$$f_2(\theta_2) + (\nu^{k+1})^\top A_2 \theta_2.$$

So we have

$$\begin{aligned} f_1(\theta_1^{k+1}) &+ \left(\nu^{k+1} - \mu A_2(\theta_2^{k+1} - \theta_2^k)\right)^\top A_1 \theta_1^{k+1} \\ &\le f_1(\theta_1^\star) + \left(\nu^{k+1} - \mu A_2(\theta_2^{k+1} - \theta_2^k)\right)^\top A_1 \theta_1^\star, \end{aligned}$$

and

$$f_2(\theta_2^{k+1}) + (\nu^{k+1})^\top A_2 \theta_2^{k+1} \le f_2(\theta_2^\star) + (\nu^{k+1})^\top A_2 \theta_2^\star.$$

Proof of $(A.2)$: combine

Adding the two inequalities from the previous frame gives

$$p^{k+1} + (\nu^{k+1})^\top (A_1\theta_1^{k+1} + A_2\theta_2^{k+1}) - \mu\big(A_2(\theta_2^{k+1} - \theta_2^k)\big)^\top A_1\theta_1^{k+1}$$

$$\leq p^\star + (\nu^{k+1})^\top (A_1\theta_1^\star + A_2\theta_2^\star) - \mu\big(A_2(\theta_2^{k+1} - \theta_2^k)\big)^\top A_1\theta_1^\star.$$

Using $A_1\theta_1^\star + A_2\theta_2^\star = b$ and $r^{k+1} = A_1\theta_1^{k+1} + A_2\theta_2^{k+1} - b$, we obtain

$$p^{k+1} - p^\star \leq -(\nu^{k+1})^\top r^{k+1} + \mu\big(A_2(\theta_2^{k+1} - \theta_2^k)\big)^\top \big(A_1\theta_1^{k+1} - A_1\theta_1^\star\big).$$

Finally,

$$A_1\theta_1^{k+1} - A_1\theta_1^\star = r^{k+1} - A_2(\theta_2^{k+1} - \theta_2^\star),$$

so

$$p^{k+1} - p^\star \leq -(\nu^{k+1})^\top r^{k+1} - \mu\big(A_2(\theta_2^{k+1} - \theta_2^k)\big)^\top \big(-r^{k+1} + A_2(\theta_2^{k+1} - \theta_2^\star)\big).$$

This is exactly $(A.2)$.

Proof of $(A.1)$: first rewriting

Adding $(A.2)$ and $(A.3)$, regrouping terms, and multiplying by 2, we get

$$\begin{aligned} 2(\nu^{k+1} - \nu^\star)^\top r^{k+1} - 2\mu\big(A_2(\theta_2^{k+1} - \theta_2^k)\big)^\top r^{k+1} & \\ + 2\mu\big(A_2(\theta_2^{k+1} - \theta_2^k)\big)^\top A_2(\theta_2^{k+1} - \theta_2^\star) \leq 0. & \end{aligned} \tag{A.9}$$

For the first term, using

$$\nu^{k+1} = \nu^k + \mu r^{k+1},$$

we obtain

$$2(\nu^{k+1} - \nu^\star)^\top r^{k+1} = \frac{1}{\mu}\Big(\|\nu^{k+1} - \nu^\star\|_2^2 - \|\nu^k - \nu^\star\|_2^2\Big) + \mu\|r^{k+1}\|_2^2. \tag{A.10}$$

Proof of $(A.1)$: second rewriting

Let

$$\Delta\theta_2^{k+1} := \theta_2^{k+1} - \theta_2^k.$$

Using

$$\theta_2^{k+1} - \theta_2^\star = \Delta\theta_2^{k+1} + (\theta_2^k - \theta_2^\star),$$

the last two terms in $(A.9)$ can be rewritten as

$$\mu\|r^{k+1} - A_2\Delta\theta_2^{k+1}\|_2^2 + \mu\Big(\|A_2(\theta_2^{k+1} - \theta_2^\star)\|_2^2 - \|A_2(\theta_2^k - \theta_2^\star)\|_2^2\Big). \tag{A.11}$$

Combining $(A.10)$ and $(A.11)$ with $(A.9)$, we obtain

$$V^k - V^{k+1} \geq \mu\|r^{k+1} - A_2(\theta_2^{k+1} - \theta_2^k)\|_2^2. \tag{A.12}$$

Proof of $(A.1)$: sign of the cross term

To strengthen $(A.12)$, we need the middle term in

$$\|r^{k+1} - A_2(\theta_2^{k+1} - \theta_2^k)\|_2^2$$

to have the correct sign. Recall that θ_2^{k+1} minimizes

$$f_2(\theta_2) + (\nu^{k+1})^\top A_2\theta_2,$$

and θ_2^k minimizes

$$f_2(\theta_2) + (\nu^k)^\top A_2\theta_2.$$

Therefore,

$$f_2(\theta_2^{k+1}) + (\nu^{k+1})^\top A_2\theta_2^{k+1} \le f_2(\theta_2^k) + (\nu^{k+1})^\top A_2\theta_2^k,$$
$$f_2(\theta_2^k) + (\nu^k)^\top A_2\theta_2^k \le f_2(\theta_2^{k+1}) + (\nu^k)^\top A_2\theta_2^{k+1}.$$

Adding them gives

$$(\nu^{k+1} - \nu^k)^\top A_2(\theta_2^{k+1} - \theta_2^k) \le 0.$$

Since

$$\nu^{k+1} - \nu^k = \mu r^{k+1},$$

we obtain

$$(r^{k+1})^\top A_2(\theta_2^{k+1} - \theta_2^k) \le 0.$$

Proof of $(A.1)$: conclusion

Using the sign result from the previous frame, the square on the right-hand side of $(A.12)$ satisfies

$$\|r^{k+1} - A_2(\theta_2^{k+1} - \theta_2^k)\|_2^2 \geq \|r^{k+1}\|_2^2 + \|A_2(\theta_2^{k+1} - \theta_2^k)\|_2^2.$$

Hence

$$V^{k+1} \leq V^k - \mu\|r^{k+1}\|_2^2 - \mu\|A_2(\theta_2^{k+1} - \theta_2^k)\|_2^2,$$

which is exactly $(A.1)$.

Convergence proof: summary

We have shown:

- from $(A.1)$, $r^k \to 0$ and $s^k \to 0$;
- from $(A.2)$ and $(A.3)$, together with $r^k \to 0$ and $A_2(\theta_2^{k+1} - \theta_2^k) \to 0$, we obtain

$$p^k \to p^\star.$$

So the ADMM iterates satisfy:

$$r^k \to 0, \qquad s^k \to 0, \qquad p^k \to p^\star.$$

This completes the appendix proof of convergence.