

Negative stepsizes also make multi-block ADMM converge

Wenzhi Gao



gwz@stanford.edu ICME, Stanford University

Abstract

This blog post¹ investigates the idea from a recent advance [SA25] from minimax optimization: by periodically using negative stepsize, vanilla gradient descent ascent can be made converge on the conventional counterexamples. [SA25] provides an explanation that negative stepsizes can provide a “slingshot” effect and introduces a second-order effect. This interesting phenomenon makes the author wonder whether the idea can be generalized beyond minimax optimization, especially whether negative stepsize allows bypassing certain counterexamples for other algorithms. In the ADMM literature, there is a well-known counterexample [CHYY16] that demonstrates divergence of multi-block ADMM on simple quadratic functions. We show that periodic negative stepsize on the ADMM dual update allows multi-block ADMM to converge on the counterexample instances.

1 Introduction and background

In a recent work [SA25] on minimax optimization, it is shown that gradient descent ascent can escape the bilinear counterexamples by employing a special stepsize rule. There are several components in the stepsize rule from [SA25], and probably the most notable one is the *negative stepsize*. For simplicity, we explain the idea in the quadratic minimization context.

Consider a toy quadratic minimization problem

$$\min_{x \in \mathbb{R}^n} f(x) := \frac{1}{2} \langle x - x^*, \Lambda (x - x^*) \rangle,$$

where $\Lambda = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$ is a positive definite diagonal matrix with $L := \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n =: \mu > 0$.

Consider the gradient descent iteration with stepsize $\alpha \geq 0$:

$$x^+ = x - \alpha \nabla f(x). \tag{1}$$

We obtain, after subtracting both sides by x^* and using $\nabla f(x) = \Lambda (x - x^*)$, that

$$x^+ - x^* = x - x^* - \alpha \nabla f(x) = x - x^* - \alpha \Lambda (x - x^*) = (I - \alpha \Lambda)(x - x^*),$$

where α is the stepsize. Since $\lambda_{\max}(\Lambda) = L$ and $\lambda_{\min}(\Lambda) = \mu$, we have

$$\|I - \alpha \Lambda\| = \left\| \begin{pmatrix} 1 - \alpha L & & \\ & \ddots & \\ & & 1 - \alpha \mu \end{pmatrix} \right\| = \max\{|1 - \alpha L|, |1 - \alpha \mu|\}.$$

If $\|I - \alpha \Lambda\| < 1$, then $\|x^+ - x^*\| \leq \|I - \alpha \Lambda\| \|x - x^*\|$ and gradient descent converges. Clearly, $\|I - \alpha \Lambda\| < 1$ if and only if $\alpha \in \left(0, \frac{2}{L}\right)$.

¹. Blog generated by $\text{\TeX}_{\text{MACS}}$

Negative stepsize. The idea of using a negative stepsize is simply reverting the direction of update in (1):

$$x^+ = x + \alpha \nabla f(x).$$

Repeating the same argument as above, we have $x^+ - x^* = (I + \alpha \Lambda)(x - x^*)$. However, we always have $\lambda_{\min}(I + \alpha \Lambda) > 1$ and a negative stepsize inevitably leads to divergence.

What if we alternate between a positive and a negative stepsize? Say we take $\alpha \geq 0$ and

$$x^{++} = x^+ + \alpha \nabla f(x^+), \quad x^+ = x - \alpha \nabla f(x).$$

If we focus on per-iteration progress, then every time a negative stepsize is taken, divergence happens. However, if we analyze two steps at the same time, things become slightly different. We can write

$$\begin{aligned} x^{++} - x^* &= (I + \alpha \Lambda)(x^+ - x^*) \\ &= (I + \alpha \Lambda)(I - \alpha \Lambda)(x - x^*) \\ &= (I - \alpha^2 \Lambda^2)(x - x^*) \end{aligned}$$

and $\|I - \alpha^2 \Lambda^2\| \leq 1$ if $\alpha \in (0, \frac{\sqrt{2}}{L})$. Therefore, gradient descent still converges if it alternates between a positive and a negative stepsize. But negative stepsize clearly does not give any benefits in the quadratic case: since $|1 - x^2| \geq |1 - x|$ when $x \in (0, \sqrt{2})$, we always have $|1 - \alpha^2 \lambda_k^2| \geq |1 - \alpha \lambda_k|$.

Although the quadratic example does not show benefits of using a negative stepsize, two insights from its analysis will be useful.

1. It is useful to analyze the progress of an algorithm for more than one iteration.

A single negative causes divergence, but when we analyze two iterations, convergence is restored,

2. Some cancellation happens when alternating between positive and negative stepsizes. In particular,

$$(I + \alpha \Lambda)(I - \alpha \Lambda) = I - \alpha \Lambda + \alpha \Lambda - \alpha^2 \Lambda^2 = I - \alpha^2 \Lambda^2.$$

Now, we will move to the formal question in this post. Since negative stepsize is known to be useful in minimax optimization and allows gradient descent-ascent to escape the counterexample,

Can we find another instance to demonstrate the usefulness of negative stepsize?

The answer, as the title suggests, is the *alternating direction method of multipliers* (ADMM).

1.1 Multi-block ADMM

Consider the constrained optimization problem

$$\begin{aligned} \min_{\{x_i\}} \quad & \sum_{i=1}^m f_i(x_i) \quad \text{Dual variable} \\ \text{subject to} \quad & \sum_{i=1}^m A_i x_i = b \quad (\lambda) \end{aligned}$$

One way of solving the problem is through alternating direction methods of multipliers (ADMM). Standard two-block ($m = 2$) version of ADMM is effective in both theory and in practice:

$$\begin{aligned} x_1^{k+1} &= \arg \min_{x_1} \left\{ f_1(x_1) + f_2(x_2^k) + \langle \lambda^k, A_1 x_1 + A_2 x_2^k - b \rangle + \frac{\rho}{2} \|A_1 x_1 + A_2 x_2^k - b\|^2 \right\} \\ x_2^{k+1} &= \arg \min_{x_2} \left\{ f_1(x_1^{k+1}) + f_2(x_2) + \langle \lambda^k, A_1 x_1^{k+1} + A_2 x_2 - b \rangle + \frac{\rho}{2} \|A_1 x_1^{k+1} + A_2 x_2 - b\|^2 \right\} \\ \lambda^{k+1} &= \lambda^k + \rho (A_1 x_1^{k+1} + A_2 x_2^{k+1} - b) \end{aligned}$$

However, when the block number $m \geq 3$, it is known that the direct extension of two-block ADMM to multi-block even cannot converge for simple linear feasibility problem [CHYY16]:

$$\begin{aligned} & \min_{x_1, x_2, x_3} && 0 \\ & \text{subject to} && \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} x_1 + \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} x_2 + \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} x_3 = 0 \end{aligned} \quad (2)$$

As **Figure 1** (left) shows, three-block ADMM diverges on this simple instance. What if we try negative stepsize? The dual ascent step $\lambda^{k+1} = \lambda^k + \rho(\sum_{i=1}^m A_i x_i^{k+1} - b)$ in ADMM looks like a good candidate for trying negative stepsize. So let's replace it by

$$\lambda^{k+1} = \lambda^k + \rho(-1)^k (\sum_{i=1}^m A_i x_i^{k+1} - b), \quad (3)$$

where we simply alternate between dual ascent and descent. Although it is a naive attempt, as **Figure 1** (right) suggests, three-block ADMM with negative stepsize does converge.

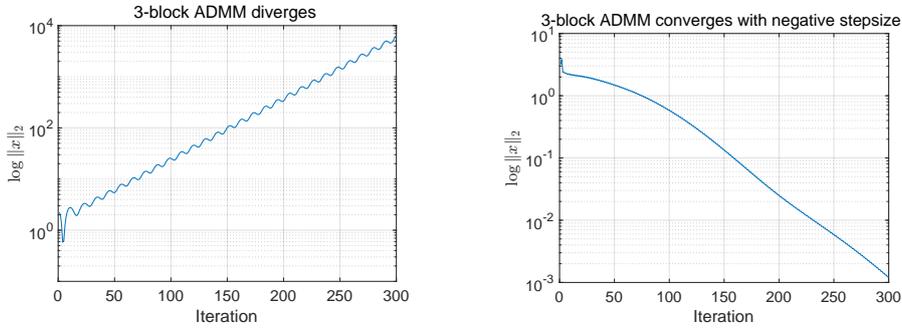


Figure 1. Left: divergence of 3-block ADMM on the counterexample. Right: convergence of 3-block ADMM with alternating dual ascent/descent stepsize.

The results so far are somewhat interesting but are fully empirical. The good news is that, the convergence behavior is not a coincidence. In fact, we can prove the following more general result:

Theorem 1. *ADMM with (alternating) negative dual stepsize (3) converges on linear feasibility problems:*

$$\min_x 0 \quad \text{subject to} \quad Ax = b,$$

where $A \in \mathbb{R}^{n \times n}$ is full-rank.

We note that several well-known attempts that address divergence of ADMM (such as random permutation [SLY20]) also start by considering this simple toy problem. For linear feasibility problem, the convergence analysis ADMM can be reduced to analyzing the spectrum of some iteration matrix. And our analysis also proceeds in this vein. We leave the technical details to **Section 2** in case it is of interest to some of the readers. But the main take-away are summarized as follows:

- Negative stepsize does find its application beyond minimax optimization.
- For this particular ADMM case, alternating between positive/negative stepsizes cancels diverging components of the spectrum of two consecutive iteration matrices.
- Unfortunately, the average-case performance of multi-block ADMM seems not improved with negative stepsize alone.

2 Analyzing ADMM with negative stepsize

This section is devoted to the proof of **Theorem 1** and we consider the following feasibility problem

$$\min_x 0 \quad \text{subject to} \quad Ax = b.$$

The proof is mostly computation. We start by explicitly writing the ADMM iteration as simple fixed-point iteration using the framework in [SLY20].

$$\begin{aligned} x_i^{k+1} &= \arg \min_{x_i} \left\{ \langle \lambda^k, \sum_{j<i} A_j x_j^{k+1} + \sum_{j>i} A_j x_j^k + A_i x_i - b \rangle + \frac{\rho}{2} \|\sum_{j<i} A_j x_j^{k+1} + \sum_{j>i} A_j x_j^k + A_i x_i - b\|^2 \right\} \\ \lambda^{k+1} &= \lambda^k + (-1)^k \rho (\sum_{i=1}^m A_i x_i^{k+1} - b) \end{aligned} \quad (4)$$

Since the objective is constant 0, without loss of generality, we can assume $\rho = 1$. (Otherwise we can scale the initial multiplier $\lambda^1 \leftarrow \lambda^1 / \rho$). The optimality condition of the subproblem gives

$$\begin{aligned} A_i^\top \lambda^k + A_i^\top (\sum_{j<i} A_j x_j^{k+1} + \sum_{j>i} A_j x_j^k + A_i x_i^{k+1} - b) &= 0 \\ \Rightarrow A_i^\top A_i x_i^{k+1} &= A_i^\top b - A_i^\top \lambda^k - A_i^\top (\sum_{j<i} A_j x_j^{k+1} + \sum_{j>i} A_j x_j^k) \end{aligned}$$

To understand the pattern of updates. Let's see what happens for $m = 3$ with $A = \begin{pmatrix} | & | & | \\ A_1 & A_2 & A_3 \\ | & | & | \end{pmatrix}$. Then

$$\begin{aligned} A_1^\top A_1 x_1^{k+1} &= A_1^\top b - A_1^\top \lambda^k - A_1^\top (A_2 x_2^k + A_3 x_3^k) \\ A_2^\top A_2 x_2^{k+1} &= A_2^\top b - A_2^\top \lambda^k - A_2^\top (A_1 x_1^{k+1} + A_3 x_3^k) \\ A_3^\top A_3 x_3^{k+1} &= A_3^\top b - A_3^\top \lambda^k - A_3^\top (A_1 x_1^{k+1} + A_2 x_2^{k+1}) \\ \lambda^{k+1} &= \lambda^k + (-1)^k (A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b), \end{aligned}$$

re-arranging the terms to move terms with index $k + 1$ to the LHS,

$$\begin{aligned} A_1^\top A_1 x_1^{k+1} &= A_1^\top b - A_1^\top \lambda^k - A_1^\top A_2 x_2^k - A_1^\top A_3 x_3^k \\ A_2^\top A_1 x_1^{k+1} + A_2^\top A_2 x_2^{k+1} &= A_2^\top b - A_2^\top \lambda^k - A_2^\top A_3 x_3^k \\ A_3^\top A_1 x_1^{k+1} + A_3^\top A_2 x_2^{k+1} + A_3^\top A_3 x_3^{k+1} &= A_3^\top b - A_3^\top \lambda^k \\ \lambda^{k+1} - (-1)^k (A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1}) &= \lambda^k - (-1)^k b, \end{aligned}$$

Hence in the matrix form, we have

$$\begin{pmatrix} \boxed{\begin{matrix} A_1^\top A_1 & 0 & 0 \\ A_2^\top A_1 & A_2^\top A_2 & 0 \\ A_3^\top A_1 & A_3^\top A_2 & A_3^\top A_3 \end{matrix}} & \begin{matrix} 0 \\ 0 \\ 0 \end{matrix} \\ A_1 & A_2 & A_3 \\ I & & \end{pmatrix} \begin{pmatrix} x_1^{k+1} \\ x_2^{k+1} \\ x_3^{k+1} \\ (-1)^{k+1} \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} A_1^\top b \\ A_2^\top b \\ A_3^\top b \\ (-1)^{k+1} b \end{pmatrix} + \begin{pmatrix} 0 & -A_1^\top A_2 & -A_1^\top A_3 & -A_1^\top \\ 0 & 0 & -A_2^\top A_3 & -A_2^\top \\ 0 & 0 & 0 & -A_3^\top \\ 0 & 0 & 0 & I \end{pmatrix} \begin{pmatrix} x_1^k \\ x_2^k \\ x_3^k \\ \lambda^k \end{pmatrix}.$$

For brevity, we define L to be the block lower-triangular part of $A^\top A$ and let $\delta := (A^\top b, -b)$, $\delta^- := (A^\top b, b)$. In the above example, we have

$$L = \begin{pmatrix} A_1^\top A_1 & 0 & 0 \\ A_2^\top A_1 & A_2^\top A_2 & 0 \\ A_3^\top A_1 & A_3^\top A_2 & A_3^\top A_3 \end{pmatrix}, \quad \delta = \begin{pmatrix} A_1^\top b \\ A_2^\top b \\ A_3^\top b \\ -b \end{pmatrix}, \quad \delta^- = \begin{pmatrix} A_1^\top b \\ A_2^\top b \\ A_3^\top b \\ b \end{pmatrix}.$$

Then the ADMM iterates can be written as linear fixed-point iterations.

Lemma 2. ([SLY20], Section 3.1). Define the state variable $z := (x_1, \dots, x_m, -\lambda)$. Then

- If k is even, then $z^{k+1} = Mz^k + L^{-1}\delta$
- If k is odd, then $z^{k+1} = M^-z^k + L^{-1}\delta^-$,

where the iteration matrices are

$$M := \begin{pmatrix} L & 0 \\ A & I \end{pmatrix}^{-1} \begin{pmatrix} L - A^\top A & A^\top \\ 0 & I \end{pmatrix} \quad \text{and} \quad M^- := \begin{pmatrix} L & 0 \\ -A & I \end{pmatrix}^{-1} \begin{pmatrix} L - A^\top A & A^\top \\ 0 & I \end{pmatrix},$$

and the matrices $\begin{pmatrix} L & 0 \\ A & I \end{pmatrix}, \begin{pmatrix} L & 0 \\ -A & I \end{pmatrix}$ are both invertible.

Given Lemma 2, if we let z, z^+, z^{++} be three consecutive steps, then we have

$$\begin{aligned} z^{++} &= M^-z^+ + L^{-1}\delta^- \\ &= M^-(Mz + L^{-1}\delta) + L^{-1}\delta^- \\ &= M^-Mz + M^-L^{-1}\delta + L^{-1}\delta^-. \end{aligned}$$

Here we can assume M^- is applied after M since it is safe to shift the index by 1.

To determine the convergence of the fixed point update

$$z^{++} = M^-Mz + M^-L^{-1}\delta + L^{-1}\delta^-,$$

it suffices to consider the spectrum of M^-M .

Lemma 3. Under the above problem setup, we have $\rho(M^-M) < 1$.

Proof. Consider

$$M^-M = \begin{pmatrix} L & 0 \\ -A & I \end{pmatrix}^{-1} \begin{pmatrix} L - A^\top A & A^\top \\ 0 & I \end{pmatrix} \begin{pmatrix} L & 0 \\ A & I \end{pmatrix}^{-1} \begin{pmatrix} L - A^\top A & A^\top \\ 0 & I \end{pmatrix}.$$

Note that $\begin{pmatrix} L & 0 \\ -A & I \end{pmatrix}^{-1} = \begin{pmatrix} L^{-1} & 0 \\ AL^{-1} & I \end{pmatrix}$ and $\begin{pmatrix} L & 0 \\ A & I \end{pmatrix}^{-1} = \begin{pmatrix} L^{-1} & 0 \\ -AL^{-1} & I \end{pmatrix}$. We have

$$\begin{aligned} M^-M &= \begin{pmatrix} L & 0 \\ -A & I \end{pmatrix}^{-1} \begin{pmatrix} L - A^\top A & A^\top \\ 0 & I \end{pmatrix} \begin{pmatrix} L & 0 \\ A & I \end{pmatrix}^{-1} \begin{pmatrix} L - A^\top A & A^\top \\ 0 & I \end{pmatrix} \\ &= \begin{pmatrix} L^{-1} & 0 \\ AL^{-1} & I \end{pmatrix} \begin{pmatrix} L - A^\top A & A^\top \\ 0 & I \end{pmatrix} \begin{pmatrix} L^{-1} & 0 \\ -AL^{-1} & I \end{pmatrix} \begin{pmatrix} L - A^\top A & A^\top \\ 0 & I \end{pmatrix}. \end{aligned}$$

We can directly multiply these matrices:

$$\begin{aligned} \begin{pmatrix} L^{-1} & 0 \\ AL^{-1} & I \end{pmatrix} \begin{pmatrix} L - A^\top A & A^\top \\ 0 & I \end{pmatrix} &= \begin{pmatrix} I - L^{-1}A^\top A & L^{-1}A^\top \\ A - AL^{-1}A^\top A & AL^{-1}A^\top + I \end{pmatrix} \\ \begin{pmatrix} L^{-1} & 0 \\ -AL^{-1} & I \end{pmatrix} \begin{pmatrix} L - A^\top A & A^\top \\ 0 & I \end{pmatrix} &= \begin{pmatrix} I - L^{-1}A^\top A & L^{-1}A^\top \\ -A + AL^{-1}A^\top A & -AL^{-1}A^\top + I \end{pmatrix}. \end{aligned}$$

Then we have

$$M^-M = \begin{pmatrix} I - L^{-1}A^\top A & L^{-1}A^\top \\ A - AL^{-1}A^\top A & AL^{-1}A^\top + I \end{pmatrix} \begin{pmatrix} I - L^{-1}A^\top A & L^{-1}A^\top \\ -A + AL^{-1}A^\top A & -AL^{-1}A^\top + I \end{pmatrix} = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix},$$

where after some tedious computation, we have

$$\begin{aligned}
M_{11} &= (I - L^{-1}A^T A)(I - L^{-1}A^T A) + L^{-1}A^T(-A + AL^{-1}A^T A) \\
&= (I - L^{-1}A^T A)(I - L^{-1}A^T A) - L^{-1}A^T A + L^{-1}A^T AL^{-1}A^T A \\
&= I - 3L^{-1}A^T A + 2L^{-1}A^T AL^{-1}A^T A \\
M_{12} &= (I - L^{-1}A^T A)L^{-1}A^T + L^{-1}A^T(-AL^{-1}A^T + I) \\
&= L^{-1}A^T - L^{-1}A^T AL^{-1}A^T - L^{-1}A^T AL^{-1}A^T + L^{-1}A^T \\
&= 2L^{-1}A^T - 2L^{-1}A^T AL^{-1}A^T \\
M_{21} &= (A - AL^{-1}A^T A)(I - L^{-1}A^T A) + (AL^{-1}A^T + I)(-A + AL^{-1}A^T A) \\
&= A - 2AL^{-1}A^T A + AL^{-1}A^T AL^{-1}A^T A - AL^{-1}A^T A \\
&\quad + AL^{-1}A^T AL^{-1}A^T A - A + AL^{-1}A^T A \\
&= -2AL^{-1}A^T A + 2AL^{-1}A^T AL^{-1}A^T A \\
M_{22} &= (A - AL^{-1}A^T A)L^{-1}A^T + (AL^{-1}A^T + I)(I - AL^{-1}A^T) \\
&= AL^{-1}A^T - AL^{-1}A^T AL^{-1}A^T + I - AL^{-1}A^T AL^{-1}A^T \\
&= I + AL^{-1}A^T - 2(AL^{-1}A^T)^2.
\end{aligned}$$

and

$$M^{-1}M = \begin{pmatrix} I - 3L^{-1}A^T A + 2L^{-1}A^T AL^{-1}A^T A & 2L^{-1}A^T - 2L^{-1}A^T AL^{-1}A^T \\ -2AL^{-1}A^T A + 2AL^{-1}A^T AL^{-1}A^T A & I + AL^{-1}A^T - 2(AL^{-1}A^T)^2 \end{pmatrix}.$$

Define $Q := L^{-1}A^T A$. We can rewrite $M^{-1}M$ as

$$M^{-1}M = \begin{pmatrix} I - 3Q + 2Q^2 & 2QA^{-1} - 2Q^2A^{-1} \\ -2AQ + 2AQ^2 & I + AQA^{-1} - 2AQ^2A^{-1} \end{pmatrix}.$$

Further notice that

$$\begin{aligned}
&\begin{pmatrix} I - 3Q + 2Q^2 & 2QA^{-1} - 2Q^2A^{-1} \\ -2AQ + 2AQ^2 & I + AQA^{-1} - 2AQ^2A^{-1} \end{pmatrix} \\
&= \begin{pmatrix} I & \\ & A \end{pmatrix} \begin{pmatrix} I - 3Q + 2Q^2 & 2Q - 2Q^2 \\ -2Q + 2Q^2 & I + Q - 2Q^2 \end{pmatrix} \begin{pmatrix} I & \\ & A^{-1} \end{pmatrix}
\end{aligned}$$

and therefore the spectral radius $M^{-1}M$ is determined by that of

$$H := \begin{pmatrix} I - 3Q + 2Q^2 & 2Q - 2Q^2 \\ -2Q + 2Q^2 & I + Q - 2Q^2 \end{pmatrix}.$$

Let $U := A^T A - L$. Note that $I - Q = I - L^{-1}(L + U) = -L^{-1}U$ is the iteration matrix of block Gauss-Seidel iteration. Hence $\rho(I - Q) \leq 1$ by the convergence of block Gauss-Seidel algorithm [ZYZL14], and it remains to connect the spectrum of Q with that of H .

Below we provide a simplified analysis for the case where Q is diagonalizable. A more careful argument can be proven in the same manner as Lemma 1 of [SLY20].

Let $Q = V \Lambda V^{-1}$ be the eigen-decomposition of Q . We have $|1 - \lambda_i| < 1$ for all i .

$$\begin{pmatrix} I - 3Q + 2Q^2 & 2Q - 2Q^2 \\ -2Q + 2Q^2 & I + Q - 2Q^2 \end{pmatrix} = \begin{pmatrix} V & \\ & V \end{pmatrix} \begin{pmatrix} I - 3\Lambda + 2\Lambda^2 & 2\Lambda - 2\Lambda^2 \\ -2\Lambda + 2\Lambda^2 & I + \Lambda - 2\Lambda^2 \end{pmatrix} \begin{pmatrix} V^{-1} & \\ & V^{-1} \end{pmatrix}$$

and it suffices to determine the spectral radius of $\begin{pmatrix} I-3\Lambda+2\Lambda^2 & 2\Lambda-2\Lambda^2 \\ -2\Lambda+2\Lambda^2 & I+\Lambda-2\Lambda^2 \end{pmatrix}$. Since we can permute the matrix to form multiple 2×2 block-diagonal matrices, it suffices to analyze

$$\rho\left(\begin{pmatrix} 1-3\lambda+2\lambda^2 & 2\lambda-2\lambda^2 \\ -2\lambda+2\lambda^2 & 1+\lambda-2\lambda^2 \end{pmatrix}\right)$$

for λ in the spectrum Λ . After direct computation, we have $\rho\left(\begin{pmatrix} 1-3\lambda+2\lambda^2 & 2\lambda-2\lambda^2 \\ -2\lambda+2\lambda^2 & 1+\lambda-2\lambda^2 \end{pmatrix}\right) = |1-\lambda|$ and

$$\rho(M^-M) = \rho(H) < 1.$$

This completes the proof. □

2.1 Proof of Theorem 1

Since ADMM can be written as fixed-point iteration

$$z^{++} = M^-Mz + M^-L^{-1}\delta + L^{-1}\delta^-$$

with $\rho(M^-M) < 1$, its convergence is immediate.

2.2 Final remarks

Recall that in our quadratic example, we mentioned that some ‘canceling’ effects happen. This is also the case for the analysis of ADMM. Note that for each of the two iteration matrices alone

$$M = \begin{pmatrix} L & 0 \\ A & I \end{pmatrix}^{-1} \begin{pmatrix} L-A^\top A & A^\top \\ 0 & I \end{pmatrix}, \quad M^- = \begin{pmatrix} L & 0 \\ -A & I \end{pmatrix}^{-1} \begin{pmatrix} L-A^\top A & A^\top \\ 0 & I \end{pmatrix},$$

we cannot guarantee $\rho(M) < 1$ or $\rho(M^-) < 1$. However, when M^-M are multiplied together, the diverging components in M, M^- cancel each other and yield $\rho(M^-M) < 1$. This is partially the ‘slingshot’ phenomenon from [SA25].

References

- [CHYY16] Caihua Chen, Bingsheng He, Yinyu Ye, and Xiaoming Yuan. The direct extension of admm for multi-block convex minimization problems is not necessarily convergent. *Mathematical Programming*, 155(1):57–79, 2016.
- [SA25] Henry Shugart and Jason M Altschuler. Negative stepsizes make gradient-descent-ascent converge. *ArXiv preprint arXiv:2505.01423*, 2025.
- [SLY20] Ruoyu Sun, Zhi-Quan Luo, and Yinyu Ye. On the efficiency of random permutation for admm and coordinate descent. *Mathematics of Operations Research*, 45(1):233–271, 2020.
- [ZYZL14] Cheng-yi Zhang, Dan Ye, Cong-Lei Zhong, and Shuanghua Luo. Convergence on gauss-seidel iterative methods for linear systems with general h-matrices. *ArXiv preprint arXiv:1410.3196*, 2014.