Blog Post #4 April 12, 2026

Non-asymptotic local convergence analysis of alternating minimization

Wenzhi Gao 

gwz@stanford.edu ICME, Stanford University

Abstract

In this blog post, we consider a proof template that allows obtaining non-asymptotic local linear convergence rate of (two-block) alternating minimization (AM). Traditional analysis for alternating minimization often treats AM as a fixed point iteration and resorts to local Jacobian linearization arguments. This type of analysis gives sharp local rates, but it is often asymptotic. We provide an alternative argument that bypasses the Jacobian argument and directly shows convergence of the original function value gap. The resulting rates are non-asymptotic but equally tight.

1 Introduction and background

Consider the convex minimization problem

$$\min_{z=(u,v)\in\mathbb{R}^{m+n}} \quad f(u,v),$$

where $f\colon \mathbb{R}^{m+n}\to\mathbb{R}$ is smooth and convex. In particular, I assume that f is twice continuously differentiable. The optimization variable $z\in\mathbb{R}^{m+n}$ is partitioned into two blocks

$$z=\begin{pmatrix} u \\ v \end{pmatrix}\in\mathbb{R}^{m+n},$$

and when partial minimization is accessible, alternating minimization is probably the simplest algorithm.

Algorithm 1. Two-block alternating minimization

input v^1
for $k=1,\dots,K$
  $u^{k+1}=\arg\min_u f(u,v^k)$
  $v^{k+1}=\arg\min_v f(u^{k+1},v)$
end for

The global convergence analysis of AM has been widely studied in the literature [BT13]. Let f be L-smooth. A standard convergence analysis of AM is to reduce the AM update to gradient descent:

Suppose $z=(u,v)$ satisfies $\nabla_v f(u,v)=0$, then $\nabla f(u,v)=(\nabla_u f(u,v),\nabla_v f(u,v))=(\nabla_u f(u,v),0)$ and

$$\begin{aligned} f(u^+,v^+)\le f(u^+,v) &= \min_u f(u,v) \\ \text{(By } u \text{ minimizes } f(\cdot,v)) \quad &\le f\Big(u-\tfrac{1}{L}\nabla_u f(u,v),v\Big) \\ \text{(By } \nabla_v f(u,v)=0) \quad &= f\Big(u-\tfrac{1}{L}\nabla_u f(u,v),v-\tfrac{1}{L}\nabla_v f(u,v)\Big) \\ \text{(By descent lemma)} \quad &\le f\Big(z-\tfrac{1}{L}\nabla f(z)\Big)\le f(z)-\tfrac{1}{2L}\|\nabla f(z)\|^2, \end{aligned}$$

and we can reduce the analysis of AM to that of gradient descent. Alternatively, since AM is invariant with respect to block-diagonal affine transformation $\begin{pmatrix} A_1 & \\ & A_1 \end{pmatrix}$, it is possible to improve the dependence on L. In either case, it is safe to believe that given any target accuracy ε, the global convergence analysis ensures that AM can reach this accuracy in

$$T^G(\varepsilon) \coloneqq \mathcal{O}\left(\mathrm{poly}\left(\frac{1}{\varepsilon}\right)\right)$$

iterations. Given this guarantee, we move on to the local story. We should not expect the global worst-case complexity T_ε^G to fully characterize the behavior of AM when $z \to z^\star$. Instead, we will resort to local analysis. The principle of local analysis is simple:

1. Analyze the algorithm on quadratics. Given a twice continuously differentiable function, it satisfies

   $$f(z) = f(z^\star) + \frac{1}{2}\|z - z^\star\|^2_{\nabla^2 f(z^\star)} + \mathcal{O}(\|z - z^\star\|^3),$$

   and analyze the same algorithm on the local approximation

   $$\lambda(z) = \lambda(u, v) = f(z^\star) + \frac{1}{2}\|z - z^\star\|^2_{\nabla^2 f(z^\star)}$$

   would provide quantitative behavior.

2. Globalize the analysis. Using a global analysis $T_\varepsilon^G \coloneqq \mathcal{O}\left(\mathrm{poly}\left(\frac{1}{\varepsilon}\right)\right)$ to ensure that the $\mathcal{O}(\|z - z^\star\|^3)$ error term is dominated by the convergence behavior on $\lambda(z)$.

The first step is often clean, while the second step is often messy and involves a lot of algebraic manipulations to control the error. So I'll only cover the first step here and leave the technical second step to **Section 2**. For AM, the local quadratic model has more structures:

$$\begin{aligned}\lambda(u, v) =& f(u^\star, v^\star) + \frac{1}{2}\left\langle \begin{pmatrix} u - u^\star \\ v - v^\star \end{pmatrix}, \begin{pmatrix} \nabla^2_{uu} f(u^\star, v^\star) & \nabla^2_{uv} f(u^\star, v^\star) \\ \nabla^2_{uv} f(u^\star, v^\star)^\top & \nabla^2_{vv} f(u^\star, v^\star) \end{pmatrix} \begin{pmatrix} u - u^\star \\ v - v^\star \end{pmatrix} \right\rangle \\ =:& f(u^\star, v^\star) + \frac{1}{2}\left\langle \begin{pmatrix} u - u^\star \\ v - v^\star \end{pmatrix}, \begin{pmatrix} P & R \\ R^\top & Q \end{pmatrix} \begin{pmatrix} u - u^\star \\ v - v^\star \end{pmatrix} \right\rangle,\end{aligned}$$

where $P \coloneqq \nabla^2_{uu} f(u^\star, v^\star), R \coloneqq \nabla^2_{uv} f(u^\star, v^\star)$, and $Q \coloneqq \nabla^2_{vv} f(u^\star, v^\star)$. Then we can explicitly write the AM update on this quadratic function

$$u^+ = \arg\min_u \lambda(u, v) = u - (u - u^\star) - P^{-1}R(v - v^\star) = u^\star - P^{-1}R(v - v^\star)$$

and similarly, $v^+ = \arg\min_v \lambda(u^+, v) = v^\star - Q^{-1}R^\top(u - u^\star)$. Note that this update is exactly preconditioned gradient descent with preconditioner $S^{-1} \coloneqq \begin{pmatrix} P & \\ & Q \end{pmatrix}^{-1}$, and we can immediately obtain

$$\begin{aligned}\|u^+ - u^\star\| &= \|P^{-1}R(v - v^\star)\| = \|P^{-1}RQ^{-1}R^\top(u - u^\star)\| \\ \|v^+ - v^\star\| &= \|Q^{-1}R^\top(u - u^\star)\| = \|Q^{-1}R^\top P^{-1}R(v - v^\star)\|.\end{aligned}$$

Hence asymptotically, the convergence of AM will be governed by the spectral radius:

$$\rho(P^{-1}RQ^{-1}R^\top) = \rho(Q^{-1}R^\top P^{-1}R) = \|P^{-1/2}RQ^{-1}R^\top P^{-1/2}\| =: \sigma_1.$$

Remark 1. Note that $\sigma_1 \in (0, 1)$ since

$$\begin{pmatrix} P & R \\ R^\top & Q \end{pmatrix} \succ 0 \Rightarrow \begin{pmatrix} P^{-1/2} & \\ & Q^{-1/2} \end{pmatrix} \begin{pmatrix} P & R \\ R^\top & Q \end{pmatrix} \begin{pmatrix} P^{-1/2} & \\ & Q^{-1/2} \end{pmatrix} = \begin{pmatrix} I & P^{-1/2}RQ^{-1/2} \\ Q^{-1/2}R^\top P^{-1/2} & I \end{pmatrix} \succ 0$$

implies that the Schur complement $I - P^{-1/2}RQ^{-1}R^\top P^{-1/2} \succ 0$.

However, to give a tight non-asymptotic analysis, we need a more explicit control. Since λ is a quadratic, the change in optimality gap can be exactly characterized:

Let u, v satisfy $\nabla\lambda_v(u,v)=0$. Then $R^\top(u-u^\star)+Q(v-v^\star)=0\Rightarrow v-v^\star=-Q^{-1}R^\top(u-u^\star)$ and

$$\begin{aligned}\lambda(u,v)-f(u^\star,v^\star)&=\tfrac{1}{2}\|u-u^\star\|_P^2+\langle u-u^\star,R(v-v^\star)\rangle+\tfrac{1}{2}\|v-v^\star\|_Q^2\\&=\tfrac{1}{2}\|u-u^\star\|_{P-RQ^{-1}R^\top}^2.\end{aligned}$$

For $u^+=u^\star-P^{-1}R(v-v^\star)$, we have

$$\lambda(u^+,v)-f(u^\star,v^\star)=\tfrac{1}{2}\|u-u^\star\|_{RQ^{-1}R^\top-RQ^{-1}R^\top P^{-1}RQ^{-1}R^\top}^2.$$

Hence it suffices to consider

$$\frac{\lambda(u^+,v)-f(u^\star,v^\star)}{\lambda(u,v)-f(u^\star,v^\star)}\le\frac{\|u-u^\star\|_{RQ^{-1}R^\top-RQ^{-1}R^\top P^{-1}RQ^{-1}R^\top}^2}{\|u-u^\star\|_{P-RQ^{-1}R^\top}^2}\le\sigma_1,$$

and the convergence behavior of AM on $\lambda(u,v)$ is expected.

After going through the messy derivation in **Section 2**, we arrive at the following convergence result.

Theorem 1. *Suppose f is L-smooth, μ-strongly convex and with H-Lipschitz Hessian. Then it takes*

$$K_\varepsilon=\left\lceil T^G\left(\min\left\{\sqrt{\tfrac{1-\sigma_1}{2}}\tfrac{\mu^3}{H^2},\sqrt[3]{\tfrac{(1-\sigma_1)^2}{16}\tfrac{3\mu^3}{8H^2}},\tfrac{\mu^3}{8H^2}\right\}\right)+\frac{2}{1-\sigma_1}\log\left(\tfrac{1}{\varepsilon}\right)\right\rceil.$$

iterations to find (u,v) such that $f(u,v)-f(u^\star,v^\star)\le\varepsilon$. Recall that T^G is a global complexity bound.

Remark 2. Although the theorem is non-asymptotic, a caveat is that it is loose in at least two aspects

- First, as I previously mentioned, AM has certain affine invariant properties that cannot be characterized by standard smoothness assumptions.
- Second, the dependence on L and μ are not ideal. L and μ are global properties of f, while the iteration trajectory of AM will always satisfy either $\nabla_u f=0$ or $\nabla_v f=0$. A tighter analysis should take this fact into account.

2 Local convergence analysis of alternating minimization

This section covers the promised AM convergence analysis. The result hinges on the following estimates.

Lemma 1. *Suppose f is L-smooth, μ-strongly convex, and has H-Lipschitz Hessian, then*

- $|f(u,v)-\lambda(u,v)|\le\frac{H}{3\mu}[f(u,v)-f(u^\star,v^\star)]^{3/2}$,
- $\|\nabla f(u,v)-\nabla\lambda(u,v)\|\le\frac{H}{\mu}[f(u,v)-f(u^\star,v^\star)]$,
- $\|\nabla^2 f(u,v)-\nabla^2\lambda(u,v)\|\le\frac{\sqrt{2}H}{\sqrt{\mu}}[f(u,v)-f(u^\star,v^\star)]^{1/2}$.

Proof. By the property of Lipschitz Hessian and strong convexity: $f(u,v)-f(u^\star,v^\star)\ge\frac{\mu}{2}\|(u-u^\star,v-v^\star)\|^2$,

$$\begin{aligned}|f(u,v)-\lambda(u,v)|&\le\tfrac{H}{6}\|(u-u^\star,v-v^\star)\|^3\le\tfrac{\sqrt{2}H}{3\mu^{3/2}}[f(u,v)-f(u^\star,v^\star)]^{3/2}\\\|\nabla f(u,v)-\nabla\lambda(u,v)\|&\le\tfrac{H}{2}\|(u-u^\star,v-v^\star)\|^2\le\tfrac{H}{\mu}[f(u,v)-f(u^\star,v^\star)]\\\|\nabla^2 f(u,v)-\nabla^2\lambda(u,v)\|&\le H\|(u-u^\star,v-v^\star)\|\le\tfrac{\sqrt{2}H}{\sqrt{\mu}}[f(u,v)-f(u^\star,v^\star)]^{1/2},\end{aligned}$$

and this completes the proof. □

The following lemma characterizes the objective value after a half alternating minimization step.

Lemma 2. *Suppose* $f(u,v)-f(u^\star,v^\star)\le\frac{\mu^3}{8H^4}$, *then*

$$f(u^+,v)-f(u^\star,v^\star)\le\sigma_1[f(u,v)-f(u^\star,v^\star)]+\frac{2\sqrt{2}H}{3\mu^{3/2}}[f(u,v)-f(u^\star,v^\star)]^{3/2}+\frac{H^2}{2\mu^3}[f(u,v)-f(u^\star,v^\star)]^2.$$

A direct application of **Lemma 2** will complete the proof.

Theorem 2. (Theorem 1) *Under the previous assumptions, it takes*

$$K_\varepsilon:=\left\lceil T^G\left(\min\left\{\sqrt{\frac{1-\sigma_1}{2}\frac{\mu^3}{H^2}},\sqrt[3]{\frac{(1-\sigma_1)^2}{16}\frac{3\mu^3}{8H^2}},\frac{\mu^3}{8H^2}\right\}\right)+\frac{1}{1-\sigma_1}\log\left(\frac{1}{\varepsilon}\right)\right\rceil$$

AM iterations to find an ε*-optimal solution.*

Proof. With **Lemma 2**, suppose $f(u,v)-f(u^\star,v^\star)\le\min\left\{\sqrt{\frac{1-\sigma_1}{2}\frac{\mu^3}{H^2}},\sqrt[3]{\frac{(1-\sigma_1)^2}{16}\frac{3\mu^3}{8H^2}}\right\}$, we have

$$\frac{2\sqrt{2}H}{3\mu^{3/2}}[f(u,v)-f(u^\star,v^\star)]^{3/2}+\frac{H^2}{2\mu^3}[f(u,v)-f(u^\star,v^\star)]^2\le\frac{1-\sigma_1}{2}[f(u,v)-f(u^\star,v^\star)]$$

and that

$$f(u^+,v)-f(u^\star,v^\star)\le\frac{1+\sigma_1}{2}[f(u,v)-f(u^\star,v^\star)].$$

Putting things together completes the proof. □

References

[BT13] Amir Beck and Luba Tetruashvili. On the convergence of block coordinate descent type methods. *SIAM journal on Optimization*, 23(4):2037–2060, 2013.

A Proof of Lemma 2

Proof. For brevity, we denote $\varepsilon:=f(u,v)-f(u^\star,v^\star)$.

Define $\alpha_v(u):=f(u,v)$ and $\ell_v(u):=\lambda(u,v)$ for fixed v. Then

$$\begin{aligned}
&f(u^+,v)-f(u^\star,v^\star)\\
=&f(u^+,v)-f(u^\star,v^\star)\\
=&\min_u\alpha_v(u)-f(u^\star,v^\star)\\
\le&\alpha_v(u-P^{-1}\nabla\alpha(u))-f(u^\star,v^\star)\\
=&\alpha_v(u-P^{-1}\nabla\alpha(u))-\ell_v(u-P^{-1}\nabla\alpha(u))+\ell_v(u-P^{-1}\nabla\alpha(u))-f(u^\star,v^\star)\\
=&\underbrace{\alpha_v(u-P^{-1}\nabla\alpha(u))-\ell_v(u-P^{-1}\nabla\alpha(u))}_{\Sigma_1}+\underbrace{\ell_v(u-P^{-1}\nabla\alpha(u))-\ell_v(u-P^{-1}\nabla\ell(u))}_{\Sigma_2}\\
&+\underbrace{\ell_v(u-P^{-1}\nabla\ell(u))-f(u^\star,v^\star)}_{\Sigma_3}.
\end{aligned}$$

and we bound $\Sigma_1,\Sigma_2,\Sigma_3$ respectively.

For Σ_1, if $\alpha(u-P^{-1}\nabla\alpha(u))\le\alpha(u)$, then $f(u-P^{-1}\nabla\alpha(u),v)\le f(u,v)$, and

$$\Sigma_1\le\frac{\sqrt{2}H}{3\mu^{3/2}}\varepsilon^{3/2}.$$

To show $\alpha(u-P^{-1}\nabla\alpha(u))\le\alpha(u)$, we first note that $\nabla^2\alpha(u)=\nabla^2_{uu}f(u,v)$ and that with

$$f(u,v)-f(u^\star,v^\star)\le\frac{\mu^3}{8H^2}\quad\Rightarrow\quad\frac{\sqrt{2}H}{\sqrt{\mu}}\sqrt{f(u,v)-f(u^\star,v^\star)}\le\frac{1}{2}\mu$$

we have

$$\begin{aligned}\nabla^2\alpha_v(u)&=\nabla^2\ell_v(u)+\nabla^2\alpha_v(u)-\nabla^2\ell_v(u)\\&\le P+\|\nabla^2\alpha_v(u)-\nabla^2\ell_v(u)\|\\&\le P+\|\nabla^2 f(u,v)-\nabla^2\lambda(u,v)\|\cdot I\\&\le P+\frac{\sqrt{2}H}{\sqrt{\mu}}[f(u,v)-f(u^\star,v^\star)]^{1/2}\cdot I\\&\le P+\frac{\mu}{2}I\le\frac{3}{2}P.\end{aligned}$$

Consider scalar function $\gamma(\theta):=\alpha_v(u-\theta P^{-1}\nabla\alpha(u))$ and let $\theta_{\max}=\sup_\theta\{\theta>0:\gamma(\theta)\le\alpha_v(u)\}$. $\theta_{\max}$ is well-defined since $P^{-1}>0$ and $P^{-1}\nabla\alpha_v(u)$ is a descent direction of α_v at u.

By convexity, we have $\alpha_v(u-\theta P^{-1}\nabla\alpha(u))\le\alpha_v(u)$ for all $\theta\le\theta_{\max}$. Then we deduce that

$$\begin{aligned}\alpha_v(u)&=\gamma(\theta_{\max})\\&=\alpha_v(u)-\theta_{\max}\langle\nabla\alpha_v(u),P^{-1}\nabla\alpha_v(u)\rangle\\&\quad+\theta^2_{\max}\textstyle\int_0^1\langle\nabla\alpha_v(u),P^{-1}\nabla^2\alpha_v(u-t\theta_{\max}P^{-1}\nabla\alpha_v(u))P^{-1}\nabla\alpha_v(u)\rangle(1-t)\,\mathrm{d}t\\&\le\alpha_v(u)-\theta_{\max}\langle\nabla\alpha_v(u),P^{-1}\nabla\alpha_v(u)\rangle+\theta^2_{\max}\textstyle\int_0^1\left\langle\nabla\alpha_v(u),P^{-1}\left(\frac{3}{2}P\right)P^{-1}\nabla\alpha_v(u)\right\rangle(1-t)\,\mathrm{d}t,\\&=\alpha_v(u)-\theta_{\max}\|\nabla\alpha_v(u)\|^2_{P^{-1}}+\frac{3}{4}\theta^2_{\max}\|\nabla\alpha_v(u)\|^2_{P^{-1}},\end{aligned}$$

where the inequality holds since $\alpha_v(u-t\theta_{\max}P^{-1}\nabla\alpha_v(u))\le\alpha_v(u)$. Given

$$\frac{3}{4}\theta^2_{\max}\|\nabla\alpha_v(u)\|^2_{P^{-1}}\ge\theta_{\max}\|\nabla\alpha_v(u)\|^2_{P^{-1}},$$

we have $\theta_{\max}\ge\frac{4}{3}$. Therefore, $\alpha_v(u-P^{-1}\nabla\alpha(u))=\gamma(1)\le\gamma(\theta_{\max})=\alpha_v(u)$.

For Σ_2, since $\ell_v(u)$ is a quadratic function, we have

$$\begin{aligned}\Sigma_2&=\ell_v(u-P^{-1}\nabla\alpha(u))-\ell_v(u-P^{-1}\nabla\ell(u))\\&=\tfrac{1}{2}\|\nabla\alpha_v(u)-\nabla\ell_v(u)\|^2_{P^{-1}}\\&\le\frac{H^2}{2\mu^3}[f(u,v)-f(u^\star,v^\star)]^2=\frac{H^2}{2\mu^3}\varepsilon^2\end{aligned}$$

For Σ_3, we have

$$\begin{aligned}\ell_v(u-P^{-1}\nabla\ell(u))-f(u^\star,v^\star)&\le\sigma_1[\ell_v(u)-f(u^\star,v^\star)]\\&=\sigma_1[\ell_v(u)-\alpha_v(u)+\alpha_v(u)-f(u^\star,v^\star)]\\&\le\sigma_1[\alpha_v(u)-f(u^\star,v^\star)]+|\ell_v(u)-\alpha_v(u)|\\&\le\sigma_1[\alpha_v(u)-f(u^\star,v^\star)]+\frac{H}{3\mu}\varepsilon^{3/2}\\&=\sigma_1\varepsilon+\frac{\sqrt{2}H}{3\mu^{3/2}}\varepsilon^{3/2}\end{aligned}$$

Putting things together completes the proof. □