Alternating Minimization

Instructor: Jin Zhang

Department of Mathematics
Southern University of Science and Technology

Fall 2023

Contents

1. The Method

Model

$$\min_{x_1 \in \mathbb{R}^{n_1}, \cdots, x_p \in \mathbb{R}^{n_p}} F(x_1, \cdots, x_p) \tag{3}$$

Notations

- A vector $x \in \prod_{i=1}^p \mathbb{R}^{n_i}$ can be written as $x = (x_1, \cdots, x_p) = (x_i)_{i=1}^p$, and define

$$\|x\| = \|(x_1, \cdots, x_p)\| = \sqrt{\sum_{i=1}^p \|x_i\|^2}$$

- For any $i = 1, \cdots, p$ we define $\mathcal{U}_i : \mathbb{R}^{n_i} \to \prod_{i=1}^p \mathbb{R}^{n_i}$ to be the linear transformation given by

$$\mathcal{U}_i(d) = (0, \cdots, 0, \underbrace{d}_{i\text{th block}}, 0, \cdots, 0), \text{ for all } d \in \mathbb{R}^{n_i}.$$

The Alternating Minimization Method

Initialization: pick $x^0 = (x_1^0, \cdots, x_p^0) \in \text{dom}(F)$.
General Step: for any $k = 0, 1, 2, \cdots$ execute the following steps:

- for $i = 1, 2, \cdots, p$, compute

$$x_i^{k+1} \in \underset{x_i \in \mathbb{R}^{n_i}}{\arg\min} F\left(x_1^{k+1}, \cdots, x_{i-1}^{k+1}, x_i, x_{i+1}^k, \cdots, x_p^k\right).$$

Notations

The kth iteration $x^k = (x_1^k, \cdots, x_p^k)$.
The kth iteration involves p subiterations, which generate the following auxiliary subsequences:

$$
\begin{aligned}
x^k = x^{k,0} &= (x_1^k, x_2^k, x_3^k, \cdots, x_p^k), \\
x^{k,1} &= (x_1^{k+1}, x_2^k, x_3^k, \cdots, x_p^k), \\
x^{k,2} &= (x_1^{k+1}, x_2^{k+1}, x_3^k, \cdots, x_p^k), \\
&\vdots \\
x^{k+1} = x^{k,p} &= (x_1^{k+1}, x_2^{k+1}, x_3^{k+1}, \cdots, x_p^{k+1}).
\end{aligned}
$$

We can alternatively rewrite the general step of the alternating minimization method as follows

The Alternating Minimization Method

Initialization: pick $x^0 = (x_1^0, \cdots, x_p^0) \in \mathrm{dom}(F)$.

General Step: for any $k = 0, 1, 2, \cdots$ execute the following steps:

- set $x^{k,0} = x^k$;
- for $i = 1, 2, \cdots, p$, compute

$$x^{k,i} = x^{k,i-1} + \mathcal{U}_i \left(\tilde{y} - x_i^k \right),$$

  where

$$\tilde{y} \in \underset{y \in \mathbb{R}^{n_i}}{\arg\min} \, F\left(x^{k,i-1} + \mathcal{U}_i \left(y - x_i^k \right) \right);$$

- set $x^{k+1} = x^{k,p}$.

Lemma (alternating minimization is well defined).

Suppose that $F : \prod_{i=1}^{p} \mathbb{R}^{n_i} \to \overline{\mathbb{R}}$ is proper and closed. Assume that F has bounded level sets; that is, $\mathrm{Lev}(F, \alpha) = \{x : F(x) \leq \alpha\}$ is bounded for any $\alpha \in \mathbb{R}$. Then F has at least one minimizer, and for any $\bar{x} \in \mathrm{dom}(F)$ and $i = 1, 2, \cdots, p$ the problem

$$\min_{y \in \mathbb{R}^{n_i}} F\left(\bar{x} + \mathcal{U}_i(y - \bar{x}_i)\right)$$

possesses a minimizer.

2. Coordinate-wise Minima

Definition (coordinate-wise minimum).

A vector $x^* \in \prod_{i=1}^{p} \mathbb{R}^{n_i}$ is a coordinate-wise minimum of $F : \prod_{i=1}^{p} \mathbb{R}^{n_i} \to \overline{\mathbb{R}}$ if $x^* \in \mathrm{dom}(F)$ and

$$F(x^*) \leq F\left(x^* + \mathcal{U}_i(y)\right) \text{ for all } i = 1, 2, \cdots, p \text{ and } y \in \mathbb{R}^{n_i}.$$

Theorem (convergence of alternating minimization to coordinate-wise minima)

Suppose that $F : \prod_{i=1}^{p} \mathbb{R}^{n_i} \to \overline{\mathbb{R}}$ is a proper closed function that is continuous over its domain. Assume that

1. for each $\bar{x} \in \text{dom}(F)$ and $i = 1, 2, \cdots, p$ the problem

$$\min_{y \in \mathbb{R}^{n_i}} F\left(\bar{x} + \mathcal{U}_i(y - \bar{x}_i)\right)$$

   has a unique minimizer.

2. the level sets of F are bounded.

Let $\left\{x^k\right\}_{k \geq 0}$ be the sequence generated by the alternating minimization method for minimizing F. Then $\left\{x^k\right\}_{k \geq 0}$ is bounded, and any limit point of the sequence is coordinate-wise minimum.

3. The Composite Model

Model

$$\min_{x_1\in\mathbb{R}^{n_1},\cdots,x_p\in\mathbb{R}^{n_p}} \left\{ F(x_1,\cdots,x_p) = f(x_1,\cdots,x_p) + \sum_{j=1}^{p} g_j(x_j) \right\}, \quad (2)$$

where $f : \prod_{j=1}^{p} \mathbb{R}^{n_j} \to \overline{\mathbb{R}}$, $g_j : \mathbb{R}^{n_j} \to \overline{\mathbb{R}}$ for $j = 1,\cdots,p$.

Notations

- The function $g : \Pi_{i=1}^{p}\mathbb{R}^{n_i} \to \overline{\mathbb{R}}$ is defined by $g(x) = \sum_{i=1}^{p} g_i(x_i)$.
- The gradient of f with respect to the ith block $(i = 1,\cdots,p)$ is denoted by $\nabla_i f$ and it holds that

$$\nabla f(x) = (\nabla_1 f(x),\cdots,\nabla_p f(x)) = (\nabla_i f(x))_{i=1}^{p}.$$

Assumption 1:

(A). $g_i : \mathbb{R}^{n_i} \to \overline{\mathbb{R}}$ is proper closed and convex for any $i = 1, \cdots, p$.

(B). $f : \Pi_{i=1}^{p} \mathbb{R}^{n_i} \to \overline{\mathbb{R}}$ is proper and closed, dom$f$ is convex, dom$g \subseteq$ int(dom$f$), and f is differentiable over int(dom$f$).

Lemma (coordinate-wise minimality⇒stationarity).

Suppose that Assumption 1 holds and that $x^* \in$ dom(g) is a coordinate-wise minimum of $F = f + g$. Then x^* is a stationary point of $F = f + g$.

Corollary

Suppose that Assumption 1 holds, and assume further that $F = f + g$ satisfies the following:

1. for each $\bar{x} \in \mathrm{dom}(F)$ and $i = 1, 2, \cdots, p$ the problem

$$\min_{y \in \mathbb{R}^{n_i}} F\left(\bar{x} + \mathcal{U}_i(y - \bar{x}_i)\right)$$

   has a unique minimizer.

2. the level sets of F are bounded.

Let $\left\{x^k\right\}_{k \geq 0}$ be the sequence generated by the alternating minimization method for minimizing $F = f + g$. Then $\left\{x^k\right\}_{k \geq 0}$ is bounded, and any limit point of the sequence is coordinate-wise minimum.

Theorem

Suppose that Assumption 1 holds and that in addition

1. f is convex;
2. f is continuous differentiable over int(dom(f));
3. the level sets of $F = f + g$ are bounded.

Then the sequence generated by the alternating minimization method for solving problem (2) is bounded, and any limit point of the sequence is an optimal solution of the problem.

Assumption 2

(A). $g_i : \mathbb{R}^{n_i} \to \overline{\mathbb{R}}$ is proper closed and convex for any $i = 1, \cdots, p$.

(B). $f : \Pi_{i=1}^{p} \mathbb{R}^{n_i} \to \overline{\mathbb{R}}$ is convex and L_f-smooth.

(C). For any $\alpha \in \mathbb{R}$, there exists $R_\alpha > 0$ such that

$$\max_{x, x^*} \{\|x - x^*\| : F(x) \leq \alpha, x^* \in X^*\} \leq R_\alpha.$$

(D). The optimal set of problem (2) is nonempty and denoted by X^*. The optimal value is denoted by F_{opt}.

Theorem ($O(1/k)$ rate of convergence of alternating minimization).

Suppose that Assumption 2 holds, and let $\{x^k\}_{k\geq 0}$ be the sequence generated by the alternating minimization method for solving problem (2). Then for $k \geq 2$,

$$F(x^k) - F_{\mathsf{opt}} \leq \max\left\{\left(\frac{1}{2}\right)^{\frac{k-1}{2}}\left(F(x^0) - F_{\mathsf{opt}}\right), \frac{8L_f p^2 R^2}{k-1}\right\},$$

where $R = R_{F(x^0)}$.

Assumption 3

(A). $g_i : \mathbb{R}^{n_i} \to \overline{\mathbb{R}}$ is proper closed and convex for $i = 1, 2$.

(B). $f : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \to \overline{\mathbb{R}}$ is convex. In addition, f is differentiable over an open set containing $\text{dom}(g)$.

(C). For $i = 1, 2$, the gradient of f is L_i-Lipschitz continuous w.r.t x_i over $\text{dom}(g_i)$.

(D). For any $\alpha \in \mathbb{R}$, there exists $R_\alpha > 0$ such that

$$\max_{x, x^*} \{\|x - x^*\| : F(x) \leq \alpha, x^* \in X^*\} \leq R_\alpha.$$

(E). The optimal set of problem (2) is nonempty and denoted by X^*. The optimal value is denoted by F_{opt}.

The Alternating Minimization Method

Initialization: pick $x_1^0 \in \mathrm{dom}(g_1), x_2^0 \in \mathrm{dom}(g_2)$ such that

$$x_2^0 \in \underset{x_2 \in \mathbb{R}^{n_2}}{\arg\min} f(x_1^0, x_2) + g_2(x_2).$$

General Step: for $k = 0, 1, \cdots,$

$$x_1^{k+1} \in \underset{x_1 \in \mathbb{R}^{n_1}}{\arg\min} f(x_1, x_2^k) + g_1(x_1),$$
$$x_2^{k+1} \in \underset{x_2 \in \mathbb{R}^{n_2}}{\arg\min} f(x_1^{k+1}, x_2) + g_2(x_2).$$

Notations

1. $x^{k+\frac{1}{2}} = \left(x_1^{k+1}, x_2^k\right)$.
2. the partial prox-grad mappings:

$$T_M^i(x) = \mathrm{prox}_{\frac{1}{M} g_i}\left(x_i - \frac{1}{M}\nabla_i f(x)\right).$$

the partial gradient mappings

$$G_M^i(x) = M\left(x - T_M^i(x)\right)$$

3. for any $M > 0$, we have

$$T_M(x) = \left(T_M^1(x), T_M^2(x)\right), \quad G_M(x) = \left(G_M^1(x), G_M^2(x)\right)$$
$$G_M^1\left(x^{k+\frac{1}{2}}\right) = G_M^2(x^k) = 0.$$

Lemma

Suppose that Assumption 3 holds. Let $\{x^k\}_{k\geq 0}$ be the sequence generated by the alternating minimization method for solving problem (2). Then for any $k \geq 0$ the following inequalities hold:

$$F(x^k) - F(x^{k+\frac{1}{2}}) \geq \frac{1}{2L_1} \left\| G^1_{L_1}(x^k) \right\|^2,$$

$$F(x^k + \frac{1}{2}) - F(x^{k+1}) \geq \frac{1}{2L_2} \left\| G^2_{L_2}(x^{k+\frac{1}{2}}) \right\|^2,$$

Lemma

Let $\{x^k\}_{k\geq 0}$ be the sequence generated by the alternating minimization method for solving problem (2). Then for any $x^* \in X^*$ and $k \geq 0$,

$$F(x^{k+\frac{1}{2}}) - F(x^*) \leq \left\| G^1_{L_1}(x^k) \right\| \cdot \left\| x^k - x^* \right\|,$$
$$F(x^{k+1}) - F(x^*) \leq \left\| G^1_{L_2}(x^{k+\frac{1}{2}}) \right\| \cdot \left\| x^{k+\frac{1}{2}} - x^* \right\|.$$

Theorem ($O(1/k)$ rate of alternating minimization—improved result)

Suppose that Assumption 3 holds, and let $\{x^k\}_{k\geq 0}$ be the sequence generated by the alternating minimization method for solving problem (2). Then for any $k \geq 2$,

$$F(x^k) - F_{\text{opt}} \leq \max\left\{\left(\frac{1}{2}\right)^{\frac{k-1}{2}}\left(F(x^0) - F_{\text{opt}}\right), \frac{8\min\{L_1, L_2\}\, p^2 R^2}{k-1}\right\},$$

where $R = R_{F(x^0)}$.