TR Global Convergence 0610

Jiaxin Hu

06/10/2020

GLOBAL CONVERGENCE PROPERTY FOR TR ALGORITHM 1

Here we study the global convergence property of iterates generated by Algorithm 1. For simplicity, let $\mathcal{A}$ denote the decision variables $(\mathcal{C}, \{M_k\})$.

Theorem 1 (Global Convergence)**.** *Assume the set* $\{\mathcal{A} | \ \mathcal{L}(\mathcal{A}) \geq \mathcal{L}(\mathcal{A}^{(0)})\}$ *is compact and the stationary points of* $\mathcal{L}(\mathcal{A})$ *are isolated module equivalence. Then any sequence* $\mathcal{A}^{(t)}$ *generated by alternating algorithm converges to a stationary point of* $\mathcal{L}(\mathcal{A})$ *up to equivalence.*

PROOF

Pick an arbitrary iterate $\mathcal{A}^{(t)}$. Because of the compactness of set $\{\mathcal{A} | \ \mathcal{L}(\mathcal{A}) \geq \mathcal{L}(\mathcal{A}^{(0)})\}$, the domain of $\mathcal{A}^{(t)}$ is bounded and thus there exists convergent sub-sequences of $\mathcal{A}^{(t)}$. Let $\mathcal{A}^*$ denote a limiting points of $\mathcal{A}^{(t)}$. Since $\mathcal{L}(\mathcal{A}^{(t)})$ increases monotonically along with $t \to \infty$, then $\mathcal{A}^*$ is a stationary point of $\mathcal{L}(\mathcal{A})$. Let $\mathcal{S} = \{\mathcal{A}^*\}$ denote the set of all the limiting points of $\mathcal{A}^{(t)}$. We have $\mathcal{S} \subset \{\mathcal{A} | \ \mathcal{L}(\mathcal{A}) \geq \mathcal{L}(\mathcal{A}^{(0)})\}$ and thus $\mathcal{S}$ is a compact set. According to [Lange, 2012], $\mathcal{S}$ is also connected.

Consider the equivalence of Tucker tensor representation. We define the equivalent class of $\mathcal{A}$ as:

$$\mathcal{E}(\mathcal{A}) = \{\mathcal{A}' | \ M_k' = M_k P_k^T, \mathcal{C}' = \mathcal{C} \times \{P_k\}, \text{ where } P_k^T \in \mathbb{O}_{r_k}, \forall k \in [K]\}.$$

Notice that, for arbitrary $\mathcal{A}$, $\mathcal{E}(\mathcal{A})$ is a non-empty open set. For arbitrary two non-equivalent points $\mathcal{A}_1$ and $\mathcal{A}_2$, we have $\mathcal{E}(\mathcal{A}_1) \bigcap \mathcal{E}(\mathcal{A}_2) = \varnothing$ and thus $\mathcal{E}(\mathcal{A}_1) \bigcup \mathcal{E}(\mathcal{A}_2)$ is not connected. Using the definition of equivalent class, let $\mathcal{S}_E$ denote the enlarged set of $\mathcal{S}$, such that:

$$\mathcal{S}_E = \bigcup_{\mathcal{A} \in \mathcal{S}} \mathcal{E}(\mathcal{A}).$$

The enlarged set $\mathcal{S}_E$ satisfies below two properties:

1. [Union of Stationary Point] The set $\mathcal{S}_E$ is an union of equivalent classes generated by the stationary points in $\mathcal{S}$.

2. [Connectedness model equivalence] The set $\mathcal{S}_E$ is connected between different equivalent classes.

Property 1 is obtained by rewriting the definition of $\mathcal{S}_E$. Property 2 is concluded by the connectedness of $\mathcal{S}$.

The isolation of stationary points and Property 1 imply that $\mathcal{S}_E$ only contains finite number of different equivalent classes. Otherwise, there is a sequence of non-equivalent stationary points whose limit is not isolated. Combined the definition of equivalent class and Property 2, we can conclude that $\mathcal{S}_E$ only contains a single equivalent class; i.e. $\mathcal{S}_E = \mathcal{E}(\mathcal{A}^*)$, where $\mathcal{A}^*$ is a stationary point of $\mathcal{L}(\mathcal{A})$. Therefore, all the convergent subsequences of $\mathcal{A}^{(t)}$ converge to one stationary point $\mathcal{A}^*$ up to equivalence.

In other words, any iterate $\mathcal{A}^{(t)}$ generated by Algorithm 1 converges to a stationary point of $\mathcal{L}(\mathcal{A})$ up to equivalence.