Tensor Ring Decomposition

Ziang Chen

Department of Mathematics
Duke University

Joint work with Yingzhou Li and Jianfeng Lu

March 24, 2021

Tensor train decomposition

- External dimension: $\vec{n} = (n_1, \ldots, n_d)$.
- Internal dimension: $\vec{r} = (r_1, \ldots, r_{d-1})$.
- $\overline{\mathcal{U}}^d_{\vec{r},\vec{n}} := \bigtimes_{i=1}^d \mathbb{R}^{r_{i-1}\times n_i \times r_i}$, where $r_0 = r_d = 1$.
- $\vec{\mathbf{u}} = (\mathbf{u}^{[1]}, \ldots, \mathbf{u}^{[d]}) \in \overline{\mathcal{U}}^d_{\vec{r},\vec{n}}$.
- $\tau : \bigtimes_{i=1}^d \mathbb{R}^{r_{i-1}\times n_i \times r_i} \longrightarrow \mathbb{R}^{n_1 \times n_2 \cdots \times n_d}$.
- $\tau(\vec{\mathbf{u}})(x_1, \ldots, x_d) = \mathbf{u}^{[1]}(x_1)\mathbf{u}^{[2]}(x_2)\cdots\mathbf{u}^{[d]}(x_d)$, where $\mathbf{u}^{[i]}(x_i) = \mathbf{u}^{[i]}(:, x_i, :)$.
- $\tau(\vec{\mathbf{u}}) = \sum_{k_1,\ldots,k_d} \mathbf{u}^{[1]}_{k_1} \otimes \mathbf{u}^{[2]}_{k_1,k_2} \otimes \cdots \otimes \mathbf{u}^{[d-1]}_{k_{d-2},k_{d-1}} \otimes \mathbf{u}^{[d]}_{k_{d-1}}$, where $\mathbf{u}^{[i]}_{k_1,k_2} := \mathbf{u}^{[i]}(k_1, :, k_2)$.

Tensor train decomposition

- Gauge greedom (gauge invariance)
- $\vec{A} = (A_1, \dots, A_{d-1})$, where $A_i \in \mathrm{GL}(r_i, \mathbb{R})$.
- $\theta_{\vec{A}}(\vec{\mathbf{u}}) = \vec{\mathbf{v}} = \left(\mathbf{v}^{[1]}, \mathbf{v}^{[2]}, \dots, \mathbf{v}^{[d]}\right)$.
- $\mathbf{v}^{[1]}(x_1) = \mathbf{u}^{[1]}(x_1)A_1$, $\mathbf{v}^{[i]}(x_i) = A_{i-1}^{-1}\mathbf{u}^{[i]}(x_i)A_i$, $\mathbf{v}^{[d]}(x_d) = A_{d-1}^{-1}\mathbf{u}^{[d]}(x_d)$.
- $\mathcal{M}_{\vec{\mathbf{u}}} = \left\{\theta_{\vec{A}}(\vec{\mathbf{u}}) \mid \vec{A} \in \times_{i=1}^{d-1} \mathrm{GL}(r_i, \mathbb{R})\right\}$.

Proposition (Rohwedder and Uschmajew, 2013)

Given $\vec{\mathbf{u}} \in \overline{\mathcal{U}}^d_{\vec{r},\vec{n}}$, the TT-rank of $\tau(\vec{\mathbf{u}})$ is $\vec{r}$ if and only if
$\tau(\vec{\mathbf{u}}) = \tau(\vec{\mathbf{v}}),\ \vec{\mathbf{v}} \in \overline{\mathcal{U}}^d_{\vec{r},\vec{n}} \iff \theta_{\vec{A}}(\vec{\mathbf{u}}) = \vec{\mathbf{v}}$ *for some* $\vec{A} \in \times_{i=1}^{d-1} \mathrm{GL}(r_i, \mathbb{R})$.

- $\min_{\vec{\mathbf{u}}\in\overline{\mathcal{U}}^d_{\vec{r},\vec{n}}} f(\vec{\mathbf{u}}) = \frac{1}{2}\|\mathbf{T} - \tau(\vec{\mathbf{u}})\|_{\mathsf{F}}^2$.

Algorithm 1 Alternating Least Square (ALS) Algorithm for TT Format

Require: Target d-th order tensor $\mathbf{T}$ and initial tensor train $\vec{\mathbf{u}}_0$.

for $\ell = 0, 1, 2, \cdots$ **do**

  for $i = 1, 2, \cdots, d$ **do**

    Perform an ALS microstep (quadratic least square problem):

$$\mathbf{u}_{\ell+1}^{[i]} = \arg\min_{\mathbf{v}} \frac{1}{2}\left\|\mathbf{T} - \tau(\mathbf{u}_{\ell+1}^{[1]}, \dots, \mathbf{u}_{\ell+1}^{[i-1]}, \mathbf{v}, \mathbf{u}_{\ell}^{[i+1]}, \dots, \mathbf{u}_{\ell}^{[d]})\right\|_{\mathsf{F}}^2$$

  end for

end for

Tensor train decomposition

Theorem (Rohwedder and Uschmajew, 2013)

Suupose that the TT-rank of $\mathbf{T}$ is $\vec{r}$. Then as long as $\vec{\mathbf{u}}_0$ is not in some measure-zero set, $\vec{\mathbf{u}}_1$ computed by ALS satisfies that $\mathbf{T} = \tau(\vec{\mathbf{u}}_1)$.

- One-sweep convergence.
- Bond dimension: $\max\{r_1, r_2, \cdots, r_{d-1}\}$.
- Consider $\vec{r} = (r, \ldots, r)$ for simplicity.
- If the bond dimension of the optimization problem is large enough, the problem can be solved by ALS trivially and the energy landscape is good enough (no (stable) spurious local minima).
- If the bond dimension of the optimization problem is strictly smaller than that of the target tensor, there may exist some spurious local minima.

Tensor ring decomposition

- External dimension: $\vec{n} = (n_1, \dots, n_d)$.
- Internal dimension: $\vec{r} = (r_1, \dots, r_d)$.
- $\overline{\mathcal{U}}^d_{\vec{r},\vec{n}} := \bigtimes_{i=1}^{d} \mathbb{R}^{r_i \times n_i \times r_{i+1}}$, where $r_{d+1} = r_1$.
- $\vec{\mathbf{u}} = (\mathbf{u}^{[1]}, \dots, \mathbf{u}^{[d]}) \in \overline{\mathcal{U}}^d_{\vec{r},\vec{n}}$.
- $\tau : \overline{\mathcal{U}}^d_{\vec{r},\vec{n}} \to \mathbb{R}^{n_1 \times n_2 \cdots \times n_d}$.
- $\tau(\vec{\mathbf{u}})(x_1, \dots, x_d) = \mathrm{tr}\big(\mathbf{u}^{[1]}(x_1)\mathbf{u}^{[2]}(x_2) \cdots \mathbf{u}^{[d]}(x_d)\big)$, where $\mathbf{u}^{[i]}(x_i) = \mathbf{u}^{[i]}(:, x_i, :)$.
- $\tau(\vec{\mathbf{u}}) = \sum_{k_1,\dots,k_d} \mathbf{u}^{[1]}_{k_1,k_2} \otimes \mathbf{u}^{[2]}_{k_2,k_3} \otimes \cdots \otimes \mathbf{u}^{[d]}_{k_d,k_{d+1}}$, where $\mathbf{u}^{[i]}_{k_1,k_2} := \mathbf{u}^{[i]}(k_1, :, k_2)$.

Tensor ring decomposition

- Gauge freedom (gauge invariant)
- $\vec{A} = (A_1, \ldots, A_d)$, where $A_i \in \mathrm{GL}(r_i, \mathbb{R})$.
- $\theta_{\vec{A}}(\vec{\mathbf{u}}) = \vec{\mathbf{v}} = \left(\mathbf{v}^{[1]}, \mathbf{v}^{[2]}, \ldots, \mathbf{v}^{[d]}\right)$.
- $\mathbf{v}^{[i]}(x_i) = A_i \mathbf{u}^{[i]}(x_i) A_{i+1}^{-1}$.
- $\mathcal{M}_{\vec{\mathbf{u}}} = \left\{\theta_{\vec{A}}(\vec{\mathbf{u}}) \mid \vec{A} \in = \times_{i=1}^{d} \mathrm{GL}(r_i, \mathbb{R})\right\}$.

Tensor ring decomposition

- Some difficulties with tensor ring format:
- TR-rank is not unique (Ye and Lim, 2018).
- It is not clear when it holds that $\tau(\vec{\mathbf{v}}) = \tau(\vec{\mathbf{u}})$ if and only if $\vec{\mathbf{v}} = \theta_{\vec{A}}(\vec{\mathbf{u}})$ for some $\vec{A} \in \times_{i=1}^{d} \mathrm{GL}(r_i, \mathbb{R})$.
- $\mathcal{R}^{d}_{\vec{r},\vec{n}} = \tau(\overline{\mathcal{U}}^{d}_{\vec{r},\vec{n}})$ is not closed (Landsberg et al., 2012).

- $\min_{\vec{\mathbf{u}} \in \overline{\mathcal{U}}^d_{\vec{r},\vec{n}}} f(\vec{\mathbf{u}}) = \frac{1}{2}\|\mathbf{T} - \tau(\vec{\mathbf{u}})\|_\mathsf{F}^2$.

Algorithm 2 Alternating Least Square (ALS) Algorithm for TR Format

Require: Target d-th order tensor $\mathbf{T}$ and initial tensor train $\vec{\mathbf{u}}_0$.

for $\ell = 0, 1, 2, \cdots$ **do**

  for $i = 1, 2, \cdots, d$ **do**

    Perform an ALS microstep (quadratic least square problem):

$$\mathbf{u}^{[i]}_{\ell+1} = \arg\min_{\mathbf{v}} \frac{1}{2}\left\|\mathbf{T} - \tau(\mathbf{u}^{[1]}_{\ell+1}, \ldots, \mathbf{u}^{[i-1]}_{\ell+1}, \mathbf{v}, \mathbf{u}^{[i+1]}_{\ell}, \ldots, \mathbf{u}^{[d]}_{\ell})\right\|_\mathsf{F}^2$$

  end for

end for

Tensor ring decomposition

- $\mathbf{T} = \tau(\vec{\mathbf{w}}) = \sum_{k_1,\ldots,k_d=1}^{r} \mathbf{w}_{k_1,k_2}^{[1]} \otimes \mathbf{w}_{k_2,k_3}^{[2]} \otimes \cdots \otimes \mathbf{w}_{k_d,k_1}^{[d]} \in \mathcal{R}_{r,\vec{n}}^{d}$.
- $\min_{\vec{\mathbf{u}} \in \overline{\mathcal{U}}_{m,\vec{n}}^{d}} \frac{1}{2} \|\mathbf{T} - \tau(\vec{\mathbf{u}})\|_{\mathrm{F}}^{2}$.
- $m = r^{d-1}$
- $\dim \ \mathrm{span} \left\{ \mathbf{w}_{k_1,k_2}^{[i]} : 1 \leqslant k_1, k_2 \leqslant r \right\} \leqslant r^2$.
- $n_i \geqslant r^2$.
- $\overline{\mathcal{W}}_{r,\vec{n}}^{d} := \left\{ \vec{\mathbf{w}} \in \overline{\mathcal{U}}_{r,\vec{n}}^{d} \mid \mathbf{w}_{k_1,k_2}^{[i]}(s) = 0, \ \forall\ 1 \leqslant i \leqslant d, \ 1 \leqslant k_1, k_2 \leqslant r, \ s \geqslant r^2 + 1 \right\}$.

Tensor ring decomposition

Theorem (C-Li-Lu, 2020)

There exists $\Omega_1 \subseteq \overline{\mathcal{W}}^d_{r,\vec{n}}$ *with* $\mu(\Omega_1) = 0$*, such that for any* $\vec{\mathbf{w}} \in \overline{\mathcal{W}}^d_{r,\vec{n}} \backslash \Omega_1$ *and* $\mathbf{T} = \tau(\vec{\mathbf{w}})$*, there exists* $\Omega_2 \subseteq \overline{\mathcal{U}}^d_{m,\vec{n}}$ *with* $\mu(\Omega_2) = 0$*, such that ALS converges to the global minimum in* d *microsteps as long as the initial point* $\vec{\mathbf{u}}_0$ *is not in* Ω_2*.*

- One-Loop Convergence.
- The method for proving that a set is of zero measure is to establish the equivalence between the set and the set of roots of a polynomial, since the measure of the root set of a non-zero polynomial is zero.

Tensor ring decomposition

- Denote $\vec{\mathbf{u}} = \vec{\mathbf{u}}_0$ and $\vec{\mathbf{v}} = \vec{\mathbf{v}}_0$.
- Suppose that we have the desired full-rank properties for $\vec{\mathbf{w}}$ and $\vec{\mathbf{u}}$.
- Set $\mathbb{X}_i := \text{span}\left\{\mathbf{w}^{[i]}_{k_1,k_2} : k_1, k_2 = 1, 2, \cdots, r\right\}$ for $i = 1, 2, \cdots, d$.
- Denote $\mathbb{P}_i$ as the orthogonal projection onto $\mathbb{X}_i$ for $i = 1, 2, \cdots, d$.
- $\frac{1}{2}\left\|\mathbf{T} - \tau(\mathbf{x}^{[1]}, \mathbf{u}^{[2]}, \ldots, \mathbf{u}^{[d]})\right\|_F^2 \geqslant \frac{1}{2}\left\|\mathbf{T} - \tau(\mathbb{P}_1(\mathbf{x}^{[1]}), \mathbf{u}^{[2]}, \ldots, \mathbf{u}^{[d]})\right\|_F^2$.
- $\mathbf{v}^{[1]}_{k_1,k_2} \in \mathbb{X}_1, \ \forall\ k_1, k_2 = 1, 2, \ldots, m.$
- $\mathbf{v}^{[i]}_{k_1,k_2} \in \mathbb{X}_i, \quad \forall k_1, k_2 = 1, 2, \ldots, m$, for $i = 1, 2, \ldots, d-1$.

Tensor Ring Decomposition: One-Loop Convergence

- $m^2 = r^{2(d-1)} \geqslant \prod_{k=1}^{d-1} \dim(\mathbb{X}_k) = \dim(\mathbb{X}_1 \otimes \cdots \otimes \mathbb{X}_{d-1})$.
- $\displaystyle\sum_{k_2,\cdots,k_{d-1}=1}^{m} \mathbf{v}_{k_1,k_2}^{[1]} \otimes \cdots \otimes \mathbf{v}_{k_{d-1},k_d}^{[d-1]} \in \mathbb{X}_1 \otimes \cdots \otimes \mathbb{X}_{d-1}$, for any $k_1, k_d = 1, \ldots, m$.
- $\operatorname{span}\left\{ \displaystyle\sum_{k_2,\ldots,k_{d-1}=1}^{m} \mathbf{v}_{k_1,k_2}^{[1]} \otimes \cdots \otimes \mathbf{v}_{k_{d-1},k_d}^{[d-1]} : k_1, k_d = 1, \ldots, m \right\} = \mathbb{X}_1 \otimes \cdots \otimes \mathbb{X}_{d-1}$.
- $\mathbf{T} \in \mathbb{X}_1 \otimes \cdots \otimes \mathbb{X}_{d-1} \otimes \mathbb{R}^{n_d}$.
- There exists $\mathbf{v}^{[d]}$ such that $\tau(\mathbf{v}^{[1]}, \ldots, \mathbf{v}^{[d-1]}, \mathbf{v}^{[d]}) = \mathbf{T}$.

Tensor ring decomposition

- $\vec{r} = r$ and $\vec{n} = n = r^2 + 1$.
- $\pi(a_1, \cdots, a_\ell) = (a_1 - 1)r^{\ell-1} + \cdots (a_{\ell-1} - 1)r + a_\ell$.
- $\mathbf{T} = \sum_{k_1,\ldots,k_d=1}^{r} \left(\bigotimes_{i=1}^{d} e_{\pi(k_{i+1},k_i)}\right) + \bigotimes_{i=1}^{d} e_n \in \mathcal{R}^d_{r+1,n}$.
- $\min_{\vec{\mathbf{u}} \in \mathcal{U}^d_{r^{d-1},n}} f(\vec{\mathbf{u}}) := \frac{1}{2}\|\mathbf{T} - \tau(\vec{\mathbf{u}})\|_F^2$

Theorem (C-Li-Lu, 2020)

For $d \geqslant 3$, $\vec{\mathbf{u}}_0 \in \mathcal{U}^d_{r^{d-1},n}$ defined via

$\vec{\mathbf{u}}_0 = (\mathbf{u}^{[1]}, \mathbf{u}^{[2]}, \cdots, \mathbf{u}^{[d]})$ *with,*

$$\mathbf{u}^{[i]}_{\pi(p_1,\ldots,p_{d-1}),\pi(q_1,\ldots,q_{d-1})} = \delta_{p_1 q_1} \cdots \delta_{p_{i-1} q_{i-1}} \delta_{p_{i+1} q_{i+1}} \cdots \delta_{p_{d-1} q_{d-1}} e_{\pi(p_i,q_i)}, \text{ and,}$$

$$\mathbf{u}^{[d]}_{\pi(p_1,\ldots,p_{d-1}),\pi(q_1,\ldots,q_{d-1})} = \delta_{p_2 q_1} \cdots \delta_{p_{d-1} q_{d-2}} e_{\pi(p_1,q_{d-1})},$$

is a local minimum of f and $f(\vec{\mathbf{u}}_0) = \frac{1}{2}$.

Tensor ring decomposition

$$\sum_{k_2,\dots,k_{d-1}=1}^{m} \mathbf{u}^{[1]}_{\pi(p_1,\dots,p_{d-1}),k_2} \otimes \mathbf{u}^{[2]}_{k_2,k_3} \otimes \cdots \otimes \mathbf{u}^{[d-2]}_{k_{d-2},k_{d-3}} \otimes \mathbf{u}^{[d-1]}_{k_{d-1},\pi(q_1,\dots,q_{d-1})}$$

$$= e_{\pi(p_1,q_1)} \otimes e_{\pi(p_2,q_2)} \otimes \cdots \otimes e_{\pi(p_{d-1},q_{d-1})}.$$

- Full-rankness.
- $\tau(\vec{\mathbf{u}}_0) = \sum_{k_1,\dots,k_d=1}^{r} \left(\bigotimes_{i=1}^{d} e_{\pi(k_{i+1},k_i)}\right)$.
- $\tau(\vec{\mathbf{u}}_0)$ it orthogonal to $\bigotimes_{i=1}^{d} e_n$.

Tensor Ring: Decomposition and "Compression"

Spurious Local Minima

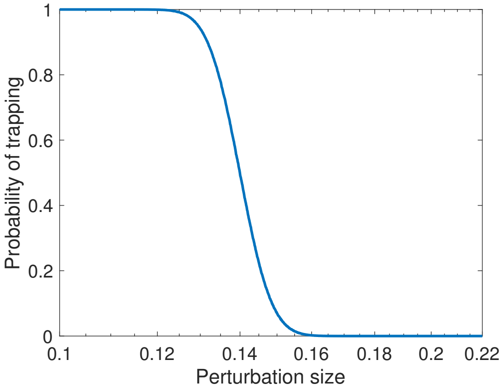

Figure: The stability of the spurious local minimum (C-Li-Lu, 2020)

References

- Chen, Z., Li, Y., & Lu, J. (2020). Tensor ring decomposition: optimization landscape and one-loop convergence of alternating least squares. SIAM Journal on Matrix Analysis and Applications, 41(3), 1416-1442.
- Landsberg, J. M., Qi, Y., & Ye, K. (2012). On the geometry of tensor network states. Quantum Information & Computation, 12(3-4), 346-354.
- Rohwedder, T., & Uschmajew, A. (2013). On local convergence of alternating schemes for optimization of convex problems in the tensor train format. SIAM Journal on Numerical Analysis, 51(2), 1134-1162.
- Ye, K., & Lim, L. H. (2018). Tensor network ranks. arXiv preprint arXiv:1801.02662.

Thanks for your attention!