Matrix Equations and Tensor Techniques
IX
Perugia, September 9-10, 2021

On the number of elements needed for low-rank tensor train completion

Stanislav Budzinskiy

Marchuk Institue of Numerical Mathematics RAS, 119333, Moscow

In this talk we explore the problem of recovering a tensor $\mathbf{A} \in \mathbb{R}^{n_1 \times \ldots \times n_d}$ with low tensor train ranks $\mathbf{r} = (1, r_1, \ldots, r_{d-1}, 1)$ from a small portion of its entries indexed by $\Omega \subset [n_1] \times \ldots \times [n_d]$:

$$\|\mathcal{R}_\Omega \mathbf{X} - \mathcal{R}_\Omega \mathbf{A}\|_F^2 \to \min \quad \text{s.t.} \quad \mathbf{X} \in \mathbb{R}^{n_1 \times \ldots \times n_d}, \quad \text{rk}_{TT}(\mathbf{X}) = \mathbf{r}.$$

By $\mathcal{R}_\Omega$ we denote the sampling operator that sets to zero all entries that are not in Ω.

In the matrix case one knows how many entries $\Omega \subset [n_1] \times [n_2]$ are needed to complete an incoherent low-rank matrix, albeit in the nuclear-norm-minimization formulation:

$$\|X\|_* \to \min \quad \text{s.t.} \quad X \in \mathbb{R}^{n_1 \times n_2}, \quad \mathcal{R}_\Omega X = \mathcal{R}_\Omega A.$$

Namely, if Ω is chosen uniformly at random with replacement then

$$O\left(r(n_1 + n_2) \log^2(n_1 + n_2)\right)$$

entries are sufficient to recover a matrix with high probability [1]. Fewer elements [2]

$$O\left(r(n_1 + n_2) \log(n_1 + n_2)\right)$$

guarantee local convergence of the Riemannian gradient descent applied to

$$\|\mathcal{R}_\Omega X - \mathcal{R}_\Omega A\|_F^2 \to \min \quad \text{s.t.} \quad X \in \mathcal{M}_r = \{X \in \mathbb{R}^{n_1 \times n_2} : \text{rk}(X) = r\}.$$

The method exploits the geometric structure of the set $\mathcal{M}_r$, which is an embedded submanifold of $\mathbb{R}^{n_1 \times n_2}$.

For tensor trains the problem is less understood (there is progress in the Tucker case [3, 4]). We follow the geometric route and establish local convergence guarantees—in terms of the sample size $|\Omega|$—of the Riemannian gradient descent for tensor train completion:

$$\|\mathcal{R}_\Omega \mathbf{X} - \mathcal{R}_\Omega \mathbf{A}\|_F^2 \to \min \quad \text{s.t.} \quad \mathbf{X} \in \mathcal{M}_\mathbf{r} = \{\mathbf{X} \in \mathbb{R}^{n_1 \times \ldots \times n_d} : \text{rk}_{TT}(\mathbf{X}) = \mathbf{r}\}.$$

To this we end we extend the notion of incoherence from matrices (their column and row subspaces) to tensor trains.

We then consider tensor completion with side information. In this problem, we are additionally given low-dimensional subspaces that contain the mode-k fiber spans of the tensor. The presence of side information makes it possible to significantly lower the number of entries sufficient for matrix completion in the nuclear norm formulation [5, 6]. We obtain similar reduction in the tensor train case.

This work has been supported by Russian Science Foundation Project (21-71-10072).

References

[1] B. Recht, *A simpler approach to matrix completion*, Journal of Machine Learning Research, (2011) pp. 3413-3430.

[2] K. Wei et al., *Guarantees of Riemannian optimization for low rank matrix completion*, arXiv:1603.06610 [math], 2016.

[3] C. Mu et al., *Square deal: Lower bounds and improved relaxations for tensor recovery*, International Conference on Machine Learning, (2014) pp. 73-81.

[4] M. Yuan, C.-H. Zhang, *On tensor completion via nuclear norm minimization*, Found Comput Math, (2016) pp. 1031-1068.

[5] P. Jain, I.S. Dhillon, *Provable inductive matrix completion*, arXiv:1306.0626 [cs, math, stat], 2013.

[6] M. Xu, R. Jin, Z. Zhou, *Speedup matrix completion with side information: Application to multi-label learning*, in Advances in Neural Information Processing Systems 26, Curran Associates Inc., 2013.