CMS 165 Foundations of Machine Learning Homework 3

An orthogonal decomposition of a symmetric tensor $T \in \mathbb{R}^{n\times n\times n}$ is a collection of orthonormal vectors $\{v_1, v_2, \dots, v_k\}$ together with corresponding positive scalars $\lambda_i > 0$ such that

$$T = \sum_{j=1}^{k} \lambda_j v_j^{\otimes 3} \tag{1}$$

In this question we will consider the variational characterization of this decomposition. First, consider the variational characterization of the eigenvalues of a symmetric matrix M. Let λ_i denote the eigenvalues and assume $\lambda_1 > \lambda_2 > \dots \lambda_k$. Then the *Rayleigh quotient*

$$\frac{u^T M u}{u^T u} \tag{2}$$

is maximized over non-zero vectors by the eigenvector associated with λ_1. Recalling the multilinear tensor notation from class, one can write the mapping as follows,

$$u \to \frac{u^T M u}{u^T u} \equiv u \to \frac{M(u,u)}{u^T u} \tag{3}$$

In tensors, the notion of eigenvectors becomes a little different than matrix case. For an extensive discussion please refer to Section 4 of [1]. For a third order tensor T, the orthogonal decomposition given in (1) is unique and $\{v_1, v_2, \dots, v_k\}$ are called robust eigenvectors. Consider the *generalized Rayleigh quotient*

$$u \to \frac{T(u,u,u)}{(u^T u)^{3/2}} \tag{4}$$

and consider the following optimization problem

$$\max_{u\in\mathbb{R}^n} T(u,u,u) \;\; s.t. \;\; \|u\| \leq 1. \tag{5}$$

Question 1. *True/False* The robust eigenvectors $\{v_1, v_2, \dots, v_k\}$ are the only stationary points of this maximization. Explain your answer in detail and compare it with matrix case.
Hint: [1] is a good resource for solving this question.

In [1], the convergence analysis of tensor power method is shown for a tensor T which has an orthogonal tensor decomposition. In this question, we will look at the case where tensors have non-orthogonal tensor decomposition. Non-orthogonal tensor decompositions can be reduced to orthogonal tensor decomposition and then tensor power method can be used to recover rank-1 components of orthogonal tensor decomposition.
Question 2. Mathematically describe the conversion of non-orthogonal tensor decomposition to orthogonal decomposition. How can we obtain the non-orthogonal components from the output of tensor power method?
Hint: In order to answer this question and for further details, it would be good to refer to TensorBook on Piazza.

Question 3. A very useful trick in non-convex optimization problems is to find hidden convexity in the problem and solve the non-convex problem by converting it to convex form. Examples can be found in [2] and [3]. Specifically in [2], Belkin et al. discussed that many important problems in machine learning can be interpreted as basis learning and demonstrated that they can be formulated as gradient ascent. Following the discussion in [2], show how the matrix eigenvector recovery and tensor orthogonal decomposition problems can be reformulated.

Question 4. You can refer to [4] to answer this problem.

4.1. Briefly explain what is a graphical model, in terms of the conditional independencies among the random variables in the model.

4.2. What is a tensor network diagram?

4.3. Refer to definitions 1.1 to 1.4 in the [4]. Explain why a graphical model associated to a graph $H = (U, \mathcal{C})$ with clique ψ_C is the same as the data of a tensor network associated to its dual graph $H^\star$ with tensors ψ_c at eah vertex of $H^\star$.

References

[1] Animashree Anandkumar, Rong Ge, Daniel Hsu, Sham M Kakade, and Matus Telgarsky. Tensor decompositions for learning latent variable models. *The Journal of Machine Learning Research*, 15(1):2773–2832, 2014.

[2] Mikhail Belkin, Luis Rademacher, and James Voss. Basis learning as an algorithmic primitive. In *Conference on Learning Theory*, pages 446–487, 2016.

[3] Yurii Nesterov and Boris T Polyak. Cubic regularization of newton method and its global performance. *Mathematical Programming*, 108(1):177–205, 2006.

[4] Elina Robeva and Anna Seigal. Duality of graphical models and tensor networks. *Information and Inference: A Journal of the IMA*, 2017.