ADDITIONAL WEB SUPPLEMENT TO "SPECTRAL AND MATRIX FACTORIZATION METHODS FOR CONSISTENT COMMUNITY DETECTION IN MULTI-LAYER NETWORKS, ANNALS OF STATISTICS, 2020"

By Subhadeep Paul * and Yuguo Chen †

The Ohio State University * *and University of Illinois at Urbana-Champaign* †

This additional web supplement contains the derivation of the algorithm for the proposed OLMF method in Section 2.1 of [2].

We start from the objective function

$$f = \text{ minimize } \sum_{m=1}^{M} \|A^{(m)} - P\Lambda^{(m)}P^T\|_F^2, \quad \text{subject to } P^TP = I.$$

We use the Lagrange method of multipliers and augment the objective function as follows:

$$\begin{aligned} O &= \sum_{m=1}^{M} \text{tr}((A^{(m)} - P\Lambda^{(m)}P^T)^T(A^{(m)} - P\Lambda^{(m)}P^T)) + \text{tr}(\Gamma_P(P^TP - I)), \\ &\equiv \sum_{m=1}^{M} \text{tr}(A^{(m)}A^{(m)} - 2\Lambda^{(m)}P^TA^{(m)}P + P\Lambda^{(m)}P^TP\Lambda^{(m)}P^T) + \text{tr}(\Gamma_P(P^TP - I)), \end{aligned}$$

where Γ_P is a symmetric matrix which is the Lagrange multiplier. To find the gradients we will use a number of results on matrix derivative of trace from section 2.4.2 of [3]. The gradient with respect to $\Lambda^{(m)}$s does not involve Γ_P and is given by

$$\frac{\partial O}{\partial \Lambda^{(m)}} : -2P^TA^{(m)}P + 2P^TP\Lambda^{(m)}P^TP.$$

The gradients with respect to P and Γ_P are given by

$$\begin{aligned} \frac{\partial O}{\partial P} &: \sum_m (-4A^{(m)}P\Lambda^{(m)} + 4P\Lambda^{(m)}P^TP\Lambda^{(m)}) + 2P\Gamma_P, \\ \frac{\partial O}{\partial \Gamma_P} &: P^TP - I. \end{aligned}$$

Setting these to 0 and solving for Γ_P we get (note setting the second gradient to 0 is equivalent to saying $P^T P = I$),

$$\Gamma_P = 2\sum_m (P^T A^{(m)} P\Lambda^{(m)} - \Lambda^{(m)}\Lambda^{(m)}).$$

and the gradient with respect to P becomes :

$$\frac{\partial O}{\partial P} : \sum_m (-4A^{(m)} P\Lambda^{(m)} + 4P\Lambda^{(m)} P^T P\Lambda^{(m)} + 4P(P^T A^{(m)} P\Lambda^{(m)} - \Lambda^{(m)}\Lambda^{(m)}))$$
$$\equiv -\sum_m (I - PP^T) A^{(m)} P\Lambda^{(m)}.$$

The same result could be obtained directly by using the formula for gradient of a function in the Grassman manifold in Section 2.5.3 in [1]. First note that P is an $n \times k$ matrix with orthonormal columns (i.e., $P^T P = I$) and therefore belongs to the Grassman manifold. In the notation of [1], we have F_Y is the derivative of the objective function with respect to P. So,

$$F_Y = \sum_m (-4A^{(m)} P\Lambda^{(m)} + 4P\Lambda^{(m)} P^T P\Lambda^{(m)})$$

Then using the formula in [1] and the fact that P is in the grassman manifold (i.e., $P^T P = I$), the gradient is

$$F_Y - YY^T F_Y$$
$$= \sum_m (-4A^{(m)} P\Lambda^{(m)} + 4P\Lambda^{(m)} P^T P\Lambda^{(m)}) - PP^T \sum_m (-4A^{(m)} P\Lambda^{(m)} + 4P\Lambda^{(m)} P^T P\Lambda^{(m)})$$
$$\equiv -\sum_m (I - PP^T) A^{(m)} P\Lambda^{(m)}$$

References.

[1] Edelman, A., Arias, T. A. and Smith, S. T. (1998). The geometry of algorithms with orthogonality constraints. *SIAM journal on Matrix Analysis and Applications* **20** 303–353.

[2] Paul, S. and Chen, Y. (2020). Spectral and matrix factorization methods for consistent community detection in multi-layer networks. *The Annals of Statistics* **48** 230–250.

[3] Petersen, K. B. and Pedersen, M. S. (2012). The Matrix Cookbook. Version 20121115.

Subhadeep Paul
Department of Statistics
The Ohio State University
Columbus, OH 43210
USA
E-mail: paul.963@osu.edu

Yuguo Chen
Department of Statistics
University of Illinois at Urbana-Champaign
Champaign, IL 61820
USA
E-mail: yuguo@illinois.edu