A sharing optimization approach for sheaf laplacian learning

Leonardo Di Nino

April 2024

1 Setting up the problem

The problem we want to solve is the following one:

$$\min_{L} \quad \mathrm{tr}(X^T L X)$$

that can be recasted as following:

$$\min_{\{B_e^T B_e\}_{e \in \mathcal{E}}} \quad \mathrm{tr}(\sum_e X^T B_e^T B_e X)$$

Leveraging trace linearity and adding the regularization term we finally have:

$$\min_{\{B_e^T B_e\}_{e \in \mathcal{E}}} \quad \sum_e \mathrm{tr}(X^T B_e^T B_e X) - \lambda \log\det(\sum_e B_e^T B_e)$$

This problem is a sharing optimization one where we are basically linking the local optimization over each possible edge to the global learning problem of the laplacian. We can leverage decomposition over the local optimization, while we have no decomposition on the regularization term. If we use the usual techniques of decoupling likelihood and regularization, we get the following problem:

$$\begin{cases} \min_{\{B_e^T B_e\}_{e \in \mathcal{E}}} \sum_e \mathrm{tr}(X^T B_e^T B_e X) - \lambda \log\det\left(\sum_e D_e^T D_e\right) \\ \qquad\qquad B_e^T B_e - D_e^T D_e = 0, \quad \forall e \in \mathcal{E} \end{cases}$$

This problem can be solved in an ADMM fashion linearly with respect to the quadratic terms: this will help us in recasting the step (2) using the usual trick in dealing with shared optimization to decrease the number of variables.

$$(B_e^T B_e)^{k+1} = \underset{B_e^T B_e}{\operatorname{argmin}} \{\operatorname{tr}(X^T B_e^T B_e X) + \frac{\rho}{2} ||B_e^T B_e - (D_e^T D_e)^k + u_e^k||^2\} \tag{1}$$

$$\{D_e^T D_e\}_{e\in\mathcal{E}}^{k+1} = \underset{\{D_e^T D_e\}_{e\in\mathcal{E}}}{\operatorname{argmin}} \{-\lambda \log\det(\sum_e D_e^T D_e) + \sum_e \frac{\rho}{2} ||(B_e^T B_e)^{k+1} - D_e^T D_e + u_e^k||^2\} \tag{2}$$

$$u_e^{k+1} = u_e^k + [(B_e^T B_e)^{k+1} - (D_e^T D_e)^{k+1}] \tag{3}$$

2 The global update

Let's focus for now on step (2) that will yield a proximal mapping step if treaten properly. Setting

$$\overline{L} = \frac{1}{E} \sum_e D_e^T D_e, \quad E = |\mathcal{E}|$$

$$a_e = (B_e^T B_e)^{k+1} + u_e^k$$

we can rewrite step (2) as following:

$$\begin{cases} \min\limits_{\{D_e^T D_e\}_{e\in\mathcal{E}}} & -\lambda \log\det(E\overline{L}) + \sum_e \frac{\rho}{2} ||D_e^T D_e - a_e||^2 \\ & \overline{L} - \frac{1}{E} \sum_e D_e^T D_e = 0 \end{cases}$$

The lagrangian of the problem is

$$\mathcal{L} = -\lambda \log\det(E\overline{L}) + \sum_e \frac{\rho}{2} ||D_e^T D_e - a_e||^2 + \nu(\overline{L} - \frac{1}{E} \sum_e D_e^T D_e)$$

So that

$$\frac{\partial \mathcal{L}}{\partial D_e^T D_e} = \rho(D_e^T D_e - a_e) - \frac{\nu}{E} = 0$$

$$D_e^T D_e = a_e + \frac{\nu}{\rho E}$$

Now we can rewrite $\overline{L}$ as

$$\overline{L} = \frac{1}{E} \sum_e D_e^T D_e = \frac{1}{E} \sum_e (a_e + \frac{\nu}{\rho E}) = \overline{a} + \frac{\nu}{\rho E}$$

so that

$$\frac{\nu}{\rho E} = \overline{L} - \overline{a}$$

Finally we have the following:

$$(D_e^T D_e)^{k+1} = a_e + \overline{L} - \overline{a} = (B_e^T B_e)^{k+1} + u_e^k + \overline{L} - (\overline{B_e^T B_e})^{k+1} - \overline{u}^k \tag{4}$$

This means that the multipliers are shared in the following way:

$$u_e^{k+1} = u_e^k + (B_e^T B_e)^{k+1} - (B_e^T B_e)^{k+1} - u_e^k - \overline{L} + (\overline{B_e^T B_e})^{k+1} + \overline{u}^k$$
$$u^{k+1} = u^k + [(\overline{B_e^T B_e})^{k+1} - \overline{L}^{k+1}]$$

Now we should highlight that plugging (4) into (2) we get

$$\overline{L}^{k+1} = \underset{\overline{L}}{\operatorname{argmin}}\{-\lambda \log\det(E\overline{L}) + \frac{\rho E}{2}||\overline{L} - (\overline{B_e^T B_e}^{k+1} + u^k)||^2\}$$

which is nothing but

$$\overline{L}^{k+1} = \operatorname{prox}_{-\frac{\lambda}{\rho E}\log\det}(\overline{B_e^T B_e}^{k+1} + u^k)$$

We can leverage the results of *H. H. Bauschke and P. L. Combettes: Convex Analysis and Monotone Operator Theory in Hilbert Spaces (2nd Edition). Springer, New York, 2017* in deriving the proximal mapping of the logdet of a symmetric positive semidefinite matrix, which $\overline{L}$ is by design. In particular, given $X \in \mathbb{R}^{n\times n}$ and its spectral decomposition $X = U^T \mathrm{Diag}(s) U$, the proximal mapping of $-\log\det(X)$ is

$$\operatorname{prox}_{-\gamma\log\det}(X) = U\ \mathrm{Diag}(z)\ U^T,\ z = \frac{1}{2}(s + \sqrt{s^2 + 4\gamma}) \tag{5}$$

3 The local updates

Up to now we have collected the following ADMM scheme:

$$(B_e^T B_e)^{k+1} = \underset{B_e^T B_e}{\operatorname{argmin}} \quad \{\operatorname{tr}(X^T B_e^T B_e X) + \frac{\rho}{2}||B_e^T B_e - (B_e^T B_e)^k - \overline{L}^k + (\overline{B_e^T B_e})^k + u^k||^2\} \tag{6}$$

$$\overline{L}^{k+1} = \operatorname{prox}_{-\frac{\lambda}{\rho E}\log\det}(\overline{B_e^T B_e}^{k+1} + u^k) \tag{7}$$

$$u^{k+1} = u^k + [(\overline{B_e^T B_e})^{k+1} - \overline{L}^{k+1}] \tag{8}$$

Let's now focus on step (6), trying to define a proper learning procedure. First of all, notice that:

$$\operatorname{tr}(X^T B_e^T B_e X) = ||B_e X||^2 = ||\mathcal{F}_{u \triangleleft e} X_u - \mathcal{F}_{v \triangleleft e} X_v||^2$$

Secondly, we can also reason similarly to the second term in the function, proposing the following decomposition:

$$\frac{\rho}{2}||B_e^T B_e - (B_e^T B_e)^k - \overline{L}^k + (\overline{B_e^T B_e})^k + \overline{u}^k||^2 = \frac{\rho}{2}[l^2_{(u,u)} + l^2_{(u,v)} + l^2_{(v,u)} + l^2_{(v,v)}]$$

where

$$\begin{aligned}
l^2_{(u,u)} &= ||\mathcal{F}^T_{u \triangleleft e}\mathcal{F}_{u \triangleleft e} - (\mathcal{F}^T_{u \triangleleft e}\mathcal{F}_{u \triangleleft e})^k - \overline{L}^k_{(u,u)} + (\overline{B_e^T B_e})^k_{(u,u)} + u^k_{(u,u)}||^2 \\
l^2_{(u,v)} &= ||-\mathcal{F}^T_{u \triangleleft e}\mathcal{F}_{v \triangleleft e} + (\mathcal{F}^T_{u \triangleleft e}\mathcal{F}_{v \triangleleft e})^k - \overline{L}^k_{(u,v)} + (\overline{B_e^T B_e})^k_{(u,v)} + u^k_{(u,v)}||^2 \\
l^2_{(v,u)} &= ||-\mathcal{F}^T_{v \triangleleft e}\mathcal{F}_{u \triangleleft e} + (\mathcal{F}^T_{u \triangleleft e}\mathcal{F}_{v \triangleleft e})^k - \overline{L}^k_{(v,u)} + (\overline{B_e^T B_e})^k_{(v,u)} + u^k_{(v,u)}||^2 \\
l^2_{(v,v)} &= ||\mathcal{F}^T_{v \triangleleft e}\mathcal{F}_{v \triangleleft e} - (\mathcal{F}^T_{v \triangleleft e}\mathcal{F}_{v \triangleleft e})^k - \overline{L}^k_{(v,v)} + (\overline{B_e^T B_e})^k_{(v,v)} + u^k_{(v,v)}||^2
\end{aligned} \tag{9}$$

We can recast the local update as a biconvex problem in the restriction maps of edge e:

$$\min_{\mathcal{F}_{u \triangleleft e}, \mathcal{F}_{v \triangleleft e}} \frac{1}{2}||\mathcal{F}_{u \triangleleft e} X_u - \mathcal{F}_{v \triangleleft e} X_v||^2 + \frac{\rho}{2}[l^2_{(u,v)} + l^2_{(u,v)} + l^2_{(v,u)} + l^2_{(v,v)}] = \min_{\mathcal{F}_{u \triangleleft e}, \mathcal{F}_{v \triangleleft e}} G(\mathcal{F}_{u \triangleleft e}, \mathcal{F}_{v \triangleleft e})$$

We can tackle it through a gradient based block-coordinate descent method:

$$\frac{\partial}{\partial \mathcal{F}_{u \triangleleft e}} G(\mathcal{F}_{u \triangleleft e}, \mathcal{F}_{v \triangleleft e}) = (\mathcal{F}_{u \triangleleft e} X_u - \mathcal{F}_{v \triangleleft e} X_v) X_u^T + \rho[\mathcal{F}_{u \triangleleft e} l_{(u,u)} - 2\mathcal{F}_{v \triangleleft e} l_{(v,u)}]$$

$$\frac{\partial}{\partial \mathcal{F}_{v \triangleleft e}} G(\mathcal{F}_{u \triangleleft e}, \mathcal{F}_{v \triangleleft e}) = -(\mathcal{F}_{u \triangleleft e} X_u - \mathcal{F}_{v \triangleleft e} X_v) X_v^T + \rho[-2\mathcal{F}_{u \triangleleft e} l_{(u,v)} + \mathcal{F}_{v \triangleleft e} l_{(v,v)}]$$