OPTIMIZATION WITH CONVERGENCE GUARANTEES

The optimization problem in DeepNT is as follows,

$$\mathcal{F} = \min_{\theta, \tilde{A}} g(\theta, \tilde{A}) + \alpha ||\tilde{A}||_1, \tag{11}$$

where $g(\theta, \tilde{A}) = \mathcal{L}_{GNN} + \gamma \mathcal{L}_M + \|M \odot (\tilde{A} - A)\|_F^2$. Then the proximal gradient algorithm with extrapolation algorithm is shown as follows:

Algorithm 2: Optimization of DeepNT

Require: S, y, ω.

Ensure: Parameters θ of DeepNT, inferred adjacency matrix $\tilde{A}$.

Initialize $\tilde{A}^{-1} = \tilde{A}^0 = 0$, $\theta^{-1} = \theta^0 = 0$.

while Stopping condition is not met **do**

  $\overline{\theta}^k \leftarrow \theta^k + (1 - \omega)(\theta^k - \theta^{k-1})$

  $\overline{A}^k \leftarrow \tilde{A}^k + (1 - \omega)(\tilde{A}^k - \tilde{A}^{k-1})$.

  $\theta^{k+1} \leftarrow \overline{\theta}^k - \omega \nabla g(\overline{\theta}^k, \overline{A}^k)$.

  $\tilde{A}^{k+1} \leftarrow \arg\min_{\tilde{A}} (1/2)\|\tilde{A} - (\overline{A}^k - \omega \nabla g(\overline{\theta}^k, \overline{A}^k))\|_F^2 + \alpha \|\tilde{A}\|_1 = \text{prox}_{\omega\alpha\|\cdot\|_1}(\overline{A}^k - \omega \nabla g(\overline{\theta}^k, \overline{A}^k))$.

  if $\lambda_2(L(\tilde{A}^{k+1})) < \epsilon$ **then**

    $\tilde{A}^{k+1} \leftarrow \tilde{A}^{k+1} + \epsilon$

  end if

end while

return θ and $\tilde{A}$.

---

where $\omega > 0$ is a learning rate, and $\text{prox}_{\lambda\|\cdot\|_1}(f) = S_\lambda(f) = \arg\min_x \left(\frac{1}{2}\|x - f\|_F^2 + \lambda\|x\|_1\right)$ is the soft-thresholding operator.

Theorem 1. *Assume $g(\theta, \tilde{A})$ is Lipschitz continuous with coefficient $l > 0$, and its gradient $\nabla g(\theta, \tilde{A})$ is Lipschitz continuous with coefficient $L > 0$. Let $\frac{1}{L} \leq \omega \leq \sqrt{\frac{L}{L+l}}$, and let $\{(\theta^k, \tilde{A}^k)\}$ be a sequence generated by Algorithm 2, then any of its limit point $(\theta^*, \tilde{A}^*)$ is a stationary point of equation 11.*

Proof. With the chosen step size satisfying $\frac{1}{L} \leq \omega \leq \sqrt{\frac{L}{L+l}}$, the proximal gradient algorithm ensures:

$$\mathcal{F}(\theta^{k+1}, \tilde{A}^{k+1}) \leq \mathcal{F}(\theta^k, \tilde{A}^k).$$

This implies that the sequence $\{\mathcal{F}(\theta^k, \tilde{A}^k)\}$ is non-increasing. Since $\mathcal{F}(\theta, \tilde{A})$ is bounded below (due to the coercivity of the ℓ_1-regularization term $\alpha\|\tilde{A}\|_1$), the sequence $\{\mathcal{F}(\theta^k, \tilde{A}^k)\}$ converges to a finite value.

The boundedness of $\mathcal{F}(\theta^k, \tilde{A}^k)$ ensures that the sequence $\{(\theta^k, \tilde{A}^k)\}$ is bounded, which means $\{(\theta^k, \tilde{A}^k)\}$ has at least one limit point $(\theta^*, \tilde{A}^*)$.

Since $\|(\theta^{k+1}, \tilde{A}^{k+1}) - (\theta^k, \tilde{A}^k)\| \to 0$, and ∇g is Lipschitz continuous, it follows that:

$$\|\nabla g(\theta^{k+1}, \tilde{A}^{k+1}) - \nabla g(\theta^k, \tilde{A}^k)\| \to 0.$$

Therefore, the gradients $\nabla g(\theta^k, \tilde{A}^k)$ converge to $\nabla g(\theta^*, \tilde{A}^*)$ as $k \to \infty$.

The update for θ in Algorithm 2 is:

$$\theta^{k+1} - \overline{\theta}^k = -\omega \nabla_\theta g(\overline{\theta}^k, \overline{A}^k).$$

As $k \to \infty$, $\overline{\theta}^k \to \theta^*$, $\theta^{k+1} - \overline{\theta}^k \to 0$ and $\nabla_\theta g(\overline{\theta}^k, \overline{A}^k) \to \nabla_\theta g(\theta^*, \tilde{A}^*)$. Then, it follows that:

$$\nabla_\theta g(\theta^*, \tilde{A}^*) = \mathbf{0}.$$

The update for $\tilde{A}$ in Algorithm 2 involves solving the proximal operator:

$$\tilde{A}^{k+1} = \arg\min_{\tilde{A}} \left(\frac{1}{2} \left\| \tilde{A} - \left(\overline{A}^k - \omega \nabla_{\tilde{A}} g(\overline{\theta}^k, \overline{A}^k) \right) \right\|_F^2 + \omega\alpha \|\tilde{A}\|_1 \right).$$

This optimization is equivalent to applying the proximal mapping:

$$\tilde{A}^{k+1} = \mathrm{prox}_{\omega\alpha\|\cdot\|_1}\left(\overline{A}^k - \omega\nabla_{\tilde{A}} g(\overline{\theta}^k, \overline{A}^k)\right),$$

where $\mathrm{prox}_{\lambda\|\cdot\|_1}(f) = S_\lambda(f)$ is the soft-thresholding operator. The proximal mapping satisfies the optimality condition:

$$\mathbf{0} \in \tilde{A}^{k+1} - \left(\overline{A}^k - \omega\nabla_{\tilde{A}} g(\overline{\theta}^k, \overline{A}^k)\right) + \omega\alpha\partial\|\tilde{A}^{k+1}\|_1.$$

Rearranging this condition gives:

$$\mathbf{0} \in \nabla_{\tilde{A}} g(\overline{\theta}^k, \overline{A}^k) + \frac{1}{\omega}(\tilde{A}^{k+1} - \overline{A}^k) + \alpha\partial\|\tilde{A}^{k+1}\|_1.$$

As $k \to \infty$, the extrapolated sequence $\overline{A}^k \to \tilde{A}^*$ and the proximal updates $\tilde{A}^{k+1} \to \tilde{A}^*$. Consequently, the term $(\tilde{A}^{k+1} - \overline{A}^k)/\omega \to \mathbf{0}$. Thus, the limit point $\tilde{A}^*$ satisfies:

$$\mathbf{0} \in \nabla_{\tilde{A}} g(\theta^*, \tilde{A}^*) + \alpha\partial\|\tilde{A}^*\|_1.$$

We conclude that $(\theta^*, \tilde{A}^*)$ is a stationary point of the optimization problem since both optimality conditions are satisfied:

$$\mathbf{0} \in \nabla_\theta g(\theta^*, \tilde{A}^*), \quad \mathbf{0} \in \nabla_{\tilde{A}} g(\theta^*, \tilde{A}^*) + \alpha\partial\|\tilde{A}^*\|_1.$$

If $\lambda_2(L(\tilde{A}^{k+1})) < \epsilon$, the algorithm adjusts $\tilde{A}^{k+1}$ to ensure connectivity. This adjustment does not violate convergence guarantees because it is a bounded perturbation that preserves the descent property.

Therefore, the sequence $\{(\theta^k, \tilde{A}^k)\}$ converges to the stationary point $(\theta^*, \tilde{A}^*)$: $\lim_{k\to\infty}(\theta^k, \tilde{A}^k) = (\theta^*, \tilde{A}^*)$. This establishes the convergence of the algorithm and completes the proof.

$\square$