

Homework: Laplacians, and (A)Synchronous Gossip for Averaging

1 Problem statement and notation

Let $G = (V, E)$ be a connected undirected graph with $|V| = n$ nodes. Each node $i \in \{1, \dots, n\}$ holds a value $x_i \in \mathbb{R}$ (the vector case $x_i \in \mathbb{R}^d$ is identical componentwise). Let $x = (x_1, \dots, x_n)^\top \in \mathbb{R}^n$ and define the average and consensus vector

$$\bar{x} \triangleq \frac{1}{n} \mathbf{1}^\top x, \quad x^* \triangleq \bar{x} \mathbf{1}.$$

We measure disagreement by the energy

$$\mathcal{E}(x) \triangleq \|x - x^*\|_2^2.$$

Gossip update (pairwise averaging). When an edge $(i, j) \in E$ activates, nodes i and j replace their values by their average:

$$x_i^+ = x_j^+ = \frac{x_i + x_j}{2}, \quad x_k^+ = x_k \quad (k \neq i, j). \quad (1)$$

Weighted edges. We allow nonnegative edge weights (rates) $\lambda_{ij} = \lambda_{ji} \geq 0$ for $(i, j) \in E$.

Submission format. Submit *one* self-contained PDF report of at most **3 pages** (including figures, excluding code listings if you place them in separate files, fontsize 11pt and margin 1.5cm). In addition, submit one code file: `gossip.py`, written in **pure PyTorch** (no external ML libraries) written with less than 300 lines of code. The header should explain how to run it. Your code should reproduce the numerical experiment(s) requested in the assignment and print the key quantities you report. Drop an email to `edouard.oyallon@cncrs.fr` with tag `[MVA] HW2` in the title.

Reproducibility. Your scripts must run end-to-end from the command line without manual intervention, set a random seed, and include all hyperparameters (e.g. d , J , learning rate, number of steps) at the top of the file or via simple CLI arguments.

2 Questions

Q1. Laplacian basics and disagreement energy.

Define the *weighted Laplacian* $\Lambda \in \mathbb{R}^{n \times n}$ by

$$\Lambda \triangleq \sum_{(i,j) \in E} \lambda_{ij} (e_i - e_j)(e_i - e_j)^\top. \quad (2)$$

- (a) Show that Λ is symmetric positive semidefinite and that $\Lambda \mathbf{1} = 0$. Show that if G is connected and $\lambda_{ij} > 0$ on every edge, then

$$\ker(\Lambda) = \text{span}\{\mathbf{1}\}.$$

Give an interpretation of λ_{ij} as an edge activation intensity. We now assume that G is connected and $\lambda_{ij} > 0$ on every edge.

(b) Prove the quadratic form identity

$$x^\top \Lambda x = \sum_{(i,j) \in E} \lambda_{ij} (x_i - x_j)^2.$$

(c) Let $0 = \lambda_1(\Lambda) < \lambda_2(\Lambda) \leq \dots \leq \lambda_n(\Lambda)$ be the eigenvalues. Prove that for all $x \in \mathbb{R}^n$,

$$x^\top \Lambda x \geq \lambda_2(\Lambda) \|x - x^*\|_2^2.$$

(d) Consider the continuous-time linear consensus dynamics $\dot{x}(t) = -\Lambda x(t)$. Show that $\bar{x}(t)$ is invariant and that if $x(0) = x$, then

$$\|x(t) - x^*\|_2 \leq e^{-\lambda_2(\Lambda)t} \|x - x^*\|_2.$$

Q2. Synchronous gossip and contraction rate.

Consider the discrete-time *synchronous* gossip process: at each step, every node performs an averaging operation with its neighbors in parallel.

(a) Write the update in matrix form for some $\alpha > 0$

$$x^{k+1} = (\text{Id} - \alpha \Lambda) x^k.$$

What is the coordinate update for x_i^{k+1} ? Under which condition on α does $x^k \rightarrow x^*$? Is there an optimal α^* ? What is the practical issue with α choice, and how can we fix it? Fix an α .

(b) For α from (a), prove a contraction bound for the disagreement energy of the form

$$\mathcal{E}(x^{k+1}) \leq (1 - \rho) \mathcal{E}(x^k),$$

and express ρ in terms of the eigenvalues of Λ .

(c) For α from (a), how many *edge activations* are performed in total before $\mathcal{E}(x^k) \leq \varepsilon$?

Q3. Asynchronous gossip in continuous time and why it can be better.

Now consider *asynchronous* gossip driven by independent Poisson edge clocks: each edge $(i, j) \in E$ rings according to a probability

$$\mathcal{P}_{ij} = 2 \frac{\lambda_{ij}}{\text{Tr } \Lambda},$$

and when it rings, update by (1).

(a) Verify that \mathcal{P} is indeed a probability distribution on E .

(b) Compute $\mathbb{E}[x^{t+1} | x^t]$. Then consider $\mathbb{E}[\mathcal{E}(x^{t+1}) | x^t]$ and derive a one-step contraction bound.

(c) Deduce an exponential decay rate in wall-clock time:

$$\mathbb{E} \|x^{t+1} - x^*\|_2^2 \leq \exp(-c \lambda_1(\Lambda) t) \|x - x^*\|_2^2,$$

for an explicit constant $c > 0$.

(d) Prove that

$$\frac{\text{Tr}(\Lambda)}{\lambda_2(\Lambda)} < |E| \frac{\lambda_n(\Lambda)}{\lambda_2(\Lambda)}.$$

- (e) How many edges are activated in total before $\mathbb{E}[\mathcal{E}(x^t)] \leq \varepsilon$? Which method is best? Is it surprising? Give an example graph where one is asymptotically strictly better than the other.

Q4. Implementation and illustration.

Setup. Implement both algorithms (synchronous and asynchronous gossip) using `gloo`. Assume $\lambda_{ij} = 1$ for all $(i, j) \in E$. Fix a global communication/computation budget parameter $\nu > 0$ (e.g. a target number of edge activations per unit of training work), and ensure that both methods are compared under the same ν . $\nu = 1$ should correspond to n computed gradients for $|E|$ communications.

- (a) Explain how ν should affect your results?
 (b) Consider the training loop

$$y^t = \text{Gossip}(x^t), \quad x_i^{t+1} = \text{Optimizer}_i(\nabla f_i(y_i^t)),$$

where x_i^t denotes the local model parameters at worker i , and f_i is the local objective at worker i . Instantiate `Gossip(·)` with (i) synchronous averaging and (ii) asynchronous pairwise gossip, and instantiate `Optimizeri(·)` with a standard optimizer (reminder: no need to introduce a new optimization objective beyond stating f_i as the local loss used in training). Train a small 3-layer MLP on MNIST using SGD, and compare synchronous vs. asynchronous gossip under the same budget $\nu \in \{1, 2, 4\}$ on the following graph topologies: ring, 2D grid, star, and complete graph. For each run, track and plot the disagreement energy $\mathcal{E}(x^t)$ as a function of samples processed, and report train accuracy against the same horizontal axes. Discuss and interpret the observed trends, emphasizing the role of the topology and of ν .