Lecture 4: The Heat Equation over Graphs

Instructors: Lorenzo Orecchia, Konstantinos Ameranis

In this lecture, we study the analogue of the heat equation over graphs. Recall that the heat equation over continuous domains takes the form:

We start by considering what the correct discretization of this PDE should be over graphs. For this purpose, it is helpful to interpret Equation **??** as arising as arising as a gradient flow of a certain energy over the space of functions. We will then obtain our discretization by computing the same gradient flow of a discretized version of the energy function.

1 Heat Equation as Gradient Flow of the Dirichlet Energy

Dirichlet Energy The Dirichlet energy of a continuously differentiable function over a domain Ω is a canonical measure of the variability or non-smoothness of the function. Formally, it is defined as:

$$\mathcal{E}(f) = \frac{1}{2}\int_{\Omega} \|\nabla f(\mathbf{x})\|_2^2 \, d\mathbf{x}$$

Based on the previous lecture, the corresponding graph analogue will be

$$\mathcal{E}_G(\mathbf{x}) = \frac{1}{2}\sum_{\{i,j\}\in E} w_{ij}(x_i - x_j)^2 = \mathbf{x}^T\mathbf{L}\mathbf{x},$$

which is indeed a measure of the variability of $\mathbf{x}$ across the edges of the graph.

Steepest Descent Directions The gradient flow of a function f starting at a point $\mathbf{x}_0$ is a continuous-time dynamics which, at each instant, moves in the direction of steepest descent of the function f. To define such a direction, we must specify a choice of norm in which we measure displacement.

Definition 1. Given a continuously differentiable function $f : U \to \mathbb{R}$ on some open set $U \subseteq \mathbb{R}^n$ and a point $\mathbf{x}_0 \in U$, the steepest descent direction $\mathbf{d} \in \mathbb{R}^n$ of f at $\mathbf{x}_0$ with respect to a general norm $\|\cdot\|$ is given by:

$$\mathbf{d} = \arg\min_{\mathbf{v}\in\mathbb{R}^n} \langle \nabla f(\mathbf{x}_0), \mathbf{v}\rangle + \frac{1}{2}\cdot\|\mathbf{v}\|^2$$

Note that, by the absolute homogeneity of norms, the coefficient $1/2$ can be replaced with $h/2$ for any $h > 0$, with the only effect of scaling the resulting direction $\mathbf{d}$ by $1/h$. In words, among all directions, the steepest direction of f is the one that minimizes a trade-off of its inner product with the gradient of f at the current point, i.e., the instantaneous decrease in function value, and the amount of squared displacement $\|\mathbf{d}\|^2$ caused by $\mathbf{d}$. We can recover closed form solutions for some common choices of $\|\cdot\|$.

- **Euclidean Norm ($\|\cdot\| = \|\cdot\|_2$)** The only critical point for the function to be minimized gives:
$$\mathbf{d} = -\nabla f(\mathbf{x}_0).$$
Hence, the steepest descent direction with respect to the Euclidean norm is simply the direction of the gradient.

- **Mahalonobis norms ($\|\cdot\| = \|\cdot\|_M \stackrel{\text{def}}{=} \sqrt{\mathbf{x}^T\mathbf{M}\mathbf{x})}, \mathbf{M} \succ 0$)** By performing the change of basis $\mathbf{z} = \mathbf{D}^{-1/2}\mathbf{d}$, the minimization problem can be reduced to the Euclidean case as follows:
$$\begin{aligned}\mathbf{D}^{-1/2}\mathbf{z} &= \arg\min\langle\nabla f(\mathbf{x}_0), \mathbf{D}^{-1/2}\mathbf{z}\rangle + \frac{1}{2}\cdot\|\mathbf{z}\| \\ &= -\mathbf{D}^{-1/2}\nabla f(\mathbf{x}_0)\end{aligned}$$
Hence, $\mathbf{d} = -\mathbf{D}^{-1}\nabla f(\mathbf{x}_0)$.

We are now ready to define the gradient flow of the function f with respect to a norm $\|\cdot\|$.

Definition 2. The

In words, the tries to reduce ... while moving the least...

Formally, a gradient flow is defined by an objective function $f : U \to \mathbb{R}$ and a choice of norm $\|\cdot\|$. The following definition applies to continuously differentiable f. The interested reader is referred to Chapter 9.2 in Evans's textbook for a more general definition that extends to non-differentiable functions.

2 Simple properties of the Laplacian

1. $\mathbf{L}_G = \mathbf{D}_G - \mathbf{A}_G = \mathbf{B}_G^T\mathbf{W}_G\mathbf{B}_G \succeq 0$

2. $\lambda_{\min}(\mathbf{L}_G) = 0$

   Additionally, it is easy to show that the all-one vector $\vec{\mathbb{1}}$ always belongs to the kernel of $\mathbf{L}_G$.
$$\mathbf{L}_G\vec{\mathbb{1}} = 0$$
   The proof is quite easy, since we only need to prove that each individual entry is zero. Remember from the last lecture how we computed the ith element of the action of the Laplacian on a vector.
$$(\mathbf{L}_G)_i = d_i\left(x_i - \frac{\sum_{j\sim i} w_{ij}x_j}{d_i}\right) = d_i\left(1 - \frac{\sum_{j\sim i} w_{ij}}{d_i}\right) = d_i\left(1 - \frac{d_i}{d_i}\right) = 0$$

3. $\lambda_2(\mathbf{L}_G) = 0$ iff G is disconnected.

   Proof. We need to prove both sides.

   $\Leftarrow$ Let $S, T \subseteq V$ be two sets of vertices s.t. $|S \cap T| = 0$ and $\nexists\{u,v\} \in E : (u \in S) \cap (v \in T)$. Let $\mathbf{x}_1 = \vec{\mathbb{1}}_S$ and $\mathbf{x}_2 = \vec{\mathbb{1}}_T$. Then, it holds that $\langle\mathbf{x}_1, \mathbf{x}_2\rangle = 0$ and $\mathbf{L}_G\mathbf{x}_1 = \mathbf{L}_G\mathbf{x}_2 = 0$.

$\Rightarrow$ If the second eigenvalue is also 0, then

$$\lambda_2(\mathbf{L}_G) = 0 \Leftrightarrow \exists \mathbf{x} \perp \vec{\mathbb{1}}, \mathbf{x} \neq 0$$

But taking the quadratic form of $\mathbf{x}$ we arrive to $\mathbf{x}^T \mathbf{L}_G \mathbf{x} = 0 \Leftrightarrow \sum_{\{i,j\} \in E} w_{ij}(x_j - x_i)^2 = 0$. Notice that $\mathbf{x}$ can be thought of as a one dimensional embedding of the nodes of G. Since $\mathbf{x} \perp \vec{\mathbb{1}}$, there are going to be both positive and negative entries. Define $S_- - \{i | x_i < 0\}$ and $S_+ = \{j | x_j \geq 0\}$.

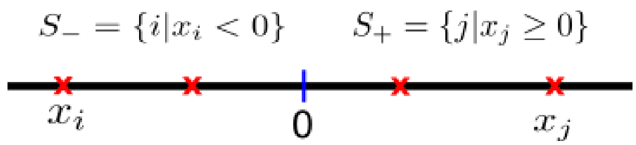

Figure 1: $\mathbf{x}$ is a 1D-embedding of the nodes of G

Suppose there was an edge $\exists \{i, j\} \in E : (i \in S_-) \cap (j \in S_+)$. Then $\mathbf{x}^T \mathbf{L}_G \mathbf{x} \geq w_{ij}(x_j - x_i)^2 > 0$ which is a contradiction. Therefore both S_- and S_+ are non empty and sets of vertices and disconnected. $\square$

3 Heat Flow as Gradient Flow

Remember that last time we saw that the Heat Equation on graphs can be expressed as follows

$$\frac{dx^{(t)}}{dt} = -D^{-1}Lx^{(t)} \qquad \frac{dx_i^{(t)}}{dt} = -\left(x_i^{(t)} = \frac{\sum_{j \sim i} w_{ij} x_j^{(t)}}{d_i} \right)$$

Gradient Flow Let there be a Lipschitz continuous function f and a vector norm $\| \cdot \|$. The gradient flow is the unit vector $\|\mathbf{v}\| = 1$ s.t. it minimizes the inner product with the gradient of f.

Definition 3. The gradient flow is the unit vector $\|\mathbf{v}\| = 1$ s.t. it minimizes the inner product with the gradient of f.

$$\mathbf{v} = \arg \min_{\|\mathbf{v}\| \leq 1} \langle \nabla f(\mathbf{x}), \mathbf{v} \rangle$$

Dual Norms We can generalize the inner product inequality to work with any norm. Specifically, given a norm $\|\cdot\|$, it holds that $\mathbf{x}^T\mathbf{y} \leq \|\mathbf{x}\| \cdot \|\mathbf{y}\|_*$. The norm $\|\cdot\|_*$ is called the dual norm. By using the previous inequality to try to define norm $\|\cdot\|_*$ we arrive to the convex problem $\|\mathbf{y}\|_* = \max_{\|\mathbf{x}\|=1} \mathbf{x}^{\mathbf{y}}$. Since the problem is convex, we can take its dual and from strong duality the optimal solutions are going to be the same. The dual problem going to also be of the form $\max_{\|\mathbf{z}\|_*=c} \mathbf{z}^T\mathbf{y}$.

Dual norms are the natural ways of defining length in dual spaces. An example of dual spaces is the vector field $\mathbf{x} \in \mathbb{R}^V$ and the flows across edges $\mathbf{y} = \mathbf{B}\mathbf{x} \in \mathbb{R}^M$.

Claim 1. *The Heat Equation is the gradient of $\mathcal{E}$ w.r.t. $\| \cdot \|_D$, i.e. it moves in the steepest descent direction of $\mathcal{E}$ w.r.t. to $\| \cdot \|_D$*

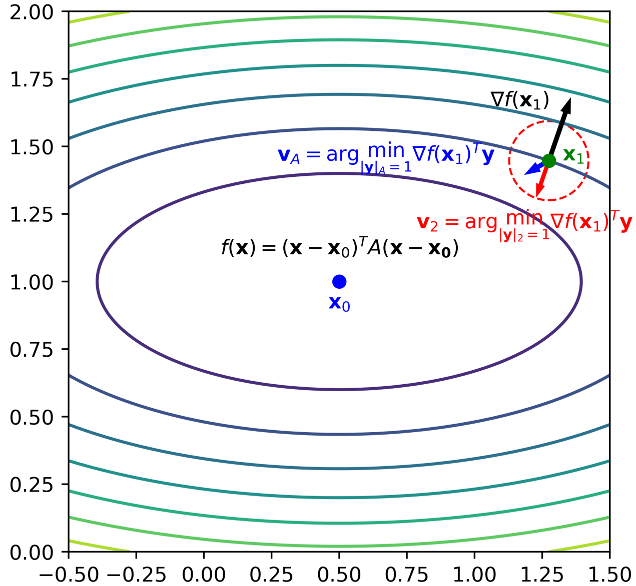

Figure 2: Steepest descent directions for function f at point $\mathbf{x}_1$ are chosen based on norm $\|\cdot\|$. For this quadratic function, you can see at point $\mathbf{x}_1$ the direction of the gradient $\nabla f(\mathbf{x}_1)$ and the steepest descent vectors $\mathbf{v}_2$ and $\mathbf{v}_A$ for which it holds $\|\mathbf{v}_2\|_2 = 1$ and $\|\mathbf{v}_A\|_A = 1$.

4 Convergence of Heat Flow

The fixed point of the Heat Equation is going to be a solution to $-\mathbf{D}^{-1}\mathbf{L}\mathbf{x} = 0 \overset{\mathbf{D}^{-1}\succ 0}{\Rightarrow} \mathbf{L}\mathbf{x} = 0$. As we saw at the start of this lecture, for connected graphs, $\mathbf{L}\mathbf{x} = 0 \Leftrightarrow \mathbf{x} \in \text{span}(\vec{\mathbb{1}})$.

The fixed point that the Heat Equation is going to converge to depends entirely by the initial $\mathbf{x}^{(0)} \in \mathbb{R}^V$. Let's take the expected value of $\mathbf{x}^{(t)}$ with weights D.

$$E_{\mathbf{D}}[x_i^{(t)}] = \sum_{i \in V} d_i x_i^{(t)} = \vec{\mathbb{1}}^T D\mathbf{x}^{(t)}$$

We want to prove that this quantity will not change as t increases.

$$\frac{dE_{\mathbf{D}}[x_i^{(t)}]}{dt} = \vec{\mathbb{1}}^T\mathbf{D}\frac{d\mathbf{x}^{(t)}}{dt} = \vec{\mathbb{1}}^T\mathbf{D}\left(-\mathbf{D}^{-1}\mathbf{L}\mathbf{x}^{(t)}\right) = -\vec{\mathbb{1}}\mathbf{L}\mathbf{x}^{(t)} = 0$$

A useful quantity for the remainder of the class is the *volume* of graph G. For a vertex $v \in V$ we define $vol(v) = d_v$ and for a set of vertices $S \subseteq V$ the volume is defined $vol(S) = \sum_{v \in S} d_v$.

Combining the previous observations we know that

$$\begin{cases} \mathbf{x}^* = c\vec{\mathbb{1}} \\ \vec{\mathbb{1}}^T\mathbf{D}\mathbf{x}^* = \vec{\mathbb{1}}\mathbf{D}\mathbf{x}_0 \end{cases} \Rightarrow c \cdot vol(V) = \vec{\mathbb{1}}^T\mathbf{D}\mathbf{x}_0 \Rightarrow c = \frac{\vec{\mathbb{1}}^T\mathbf{D}\mathbf{x}_0}{vol(V)}$$

Next Time By examining the time derivative of $\frac{d}{dt} \ln \left(\mathcal{E}(\mathbf{x}^{(t)}\right)$ we will arrive to the result that it is bounded by $\ln \left(\mathcal{E}(\mathbf{x}^{(t)})\right) - t\lambda_2(G)$.