# Lecture 5: Continuous Time Convergence of Heat Diffusion

Instructors: Lorenzo Orecchia, Konstantinos Ameranis

**Last Time** In our previous lecture, we delved into the fixed points of the Heat Equation:

$$\frac{d\mathbf{x}(t)}{dt} = -\mathbf{D}^{-1}\mathbf{L}\mathbf{x}(t)$$

We also covered the concept of gradient flow, which entails moving in the steepest descent direction w.r.t. a norm $\|\cdot\|$. To elucidate this, we defined the Dirichlet energy of a graph, where $\mathbf{x} \in \mathbb{R}^V$ defines a function on the nodes as

$$\mathcal{U}(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{L}\mathbf{x}$$

we proved that the time derivative of the solution to the Heat Equation is simply the gradient flow of the Dirichlet energy of the graph.

## 1 Convergence of Heat Equation

While we have introduced the Heat Equation, we haven't yet substantiated much about its solutions. In the previous session, we illustrated that $\vec{\mathbb{1}}\mathbf{D}\mathbf{x}(t) = \vec{\mathbb{1}}\mathbf{D}\mathbf{x}_0$and, in cases of a connected graph, a unique fixed point $\bar{x}\vec{\mathbb{1}} = \frac{\vec{\mathbb{1}}\mathbf{D}\mathbf{x}_0}{\mathrm{vol}(V)}\vec{\mathbb{1}}$ exists.

**Idea:** If we pick a distance measure $d$ from the fixed point $\mathbf{x}^* = \bar{x}\vec{\mathbb{1}}$ and show that

$$\lim_{t\to\infty} d\left(\mathbf{x}(t), \mathbf{x}^*\right) = 0 \Leftrightarrow \lim_{t\to\infty} \mathbf{x}(t) = \mathbf{x}^*$$

we will have proved convergence, since all metrics require $d(x, y) = 0 \Leftrightarrow x = y$.

**Notation:** Going forward, we will frequently employ the quantity $\pi_i = \frac{d_i}{\mathrm{vol}(V)}$ so it is important to introduce it early. This quantity is the stationary distribution for the natural random walk.

The distance measure[1] that will make this analysis work is

$$d\left(\mathbf{x}(t), \mathbf{x}^*\right) = \|\mathbf{x}(t) - \mathbf{x}^*\|_{\mathbf{D}}^2$$

A more natural way to think of this is as the variation

$$\operatorname*{Var}_{i\sim\pi_i}\left(\mathbf{x}(t)\right) = \sum_{i\in V}\pi_i(x_i - \bar{x})^2 = \frac{1}{\mathrm{Vol}(V)}\|\mathbf{x}(t) - \bar{x}\vec{\mathbb{1}}\|_{\mathbf{D}}^2$$

Notice that the two quantities are equal up to a constant.

[1]The chosen distance measure is also called a Lyapunov function

## 1.1 Change in distance

The difference between the final and starting distance is the integral of the time derivative.

$$d\left(\mathbf{x}(t), \bar{x}\vec{\mathbb{1}}\right) - d\left(\mathbf{x}(0), \bar{x}\vec{\mathbb{1}}\right) = \int_0^T \frac{d}{dt} d\left(\mathbf{x}(\tau), \bar{x}\vec{\mathbb{1}}\right) d\tau$$

Our plan is first to compute that derivative.

$$\begin{aligned}
\frac{d}{dt} d\left(\mathbf{x}(t), \bar{x}\vec{\mathbb{1}}\right) &= \frac{d}{dt}\frac{1}{2}\|\mathbf{x}(t) - \bar{x}\vec{\mathbb{1}}\|_{\mathbf{D}}^2 \\
&= \langle \mathbf{x}(t) - \bar{x}\vec{\mathbb{1}}, \frac{d\mathbf{x}(t)}{dt}\rangle_{\mathbf{D}} \\
&= \langle \mathbf{x}(t) - \bar{x}\vec{\mathbb{1}}, -\mathbf{D}^{-1}\mathbf{L}\mathbf{x}(t)\rangle_{\mathbf{D}} \\
&= -\langle \mathbf{x}(t) - \bar{x}\vec{\mathbb{1}}, \mathbf{L}\mathbf{x}(t)\rangle \\
&= -\mathbf{x}(t)^T\mathbf{L}\mathbf{x}(t) + \bar{x}\vec{\mathbb{1}}\mathbf{L}\mathbf{x}(t) \\
&= -\mathbf{x}(t)^T\mathbf{L}\mathbf{x}(t)
\end{aligned}$$

Instead of tracking the distance, we will instead track the natural logarithm of the distance, as it better showcases the convergence rate we aim to achieve.

$$\frac{d}{dt}\ln\left(d\left(\mathbf{x}(t), \bar{x}\vec{\mathbb{1}}\right)\right) = -2\frac{\mathbf{x}(t)^T\mathbf{L}\mathbf{x}(t)}{\|\mathbf{x}(t) - \bar{x}\vec{\mathbb{1}}\|_{\mathbf{D}}^2} = -2\frac{\mathbf{x}(t)^T\mathbf{L}\mathbf{x}(t)}{\mathbf{x}(t)^T\mathbf{D}\mathbf{x}(t)} \leq \underbrace{-2\min_{\mathbf{x}\perp\vec{\mathbb{1}}}\frac{\mathbf{x}^T\mathbf{L}\mathbf{x}}{\mathbf{x}^T\mathbf{D}\mathbf{x}}}_{\text{Worst case}}$$

## 1.2 Generalized Eigenvalues/vectors

Standard eigenvalues/vectors are defined through the Courant-Fisher theorem as

$$\lambda_k(\mathbf{L}) = \max_{\substack{S\subseteq\mathbb{R}^n \\ \dim(S)=k}} \min_{\substack{\mathbf{x}\in S \\ \mathbf{x}\neq 0}} \frac{\mathbf{x}^T\mathbf{L}\mathbf{x}}{\mathbf{x}^T\mathbf{x}}$$

Taking optimality conditions we already saw that the solution is of the form $\mathbf{L}\mathbf{x} = \lambda\mathbf{x}$. This definition uses the standard inner product $\langle\cdot,\cdot\rangle_{\mathbf{I}}$. The generalized eigenvectors/values are defined w.r.t. to the inner product $\langle\cdot,\cdot\rangle_D$ and through the Courant-Fisher theorem they are the solution to

$$\lambda_k(\mathbf{L},\mathbf{D}) = \max_{\substack{S\subseteq\mathbb{R}^n \\ \dim(S)=k}} \min_{\substack{\mathbf{x}\in S \\ \mathbf{x}\neq 0}} \frac{\mathbf{x}^T\mathbf{L}\mathbf{x}}{\mathbf{x}^T\mathbf{D}\mathbf{x}}$$

Repeating the optimality conditions it is proven that the eigenvalues will be of the form $\mathbf{L}\mathbf{x} = \lambda\mathbf{D}\mathbf{x}$.

Another way to prove the above result is through a change of variables that will allow us to find the standard eigenvalues/vectors w.r.t. to $\langle\cdot,\cdot\rangle_{\mathbf{I}}$. To achieve that we need to set $\mathbf{y} = \mathbf{D}^{1/2}\mathbf{x}$. In that case, $\mathbf{L}\mathbf{D}^{-1/2}\mathbf{y} = \lambda\mathbf{D}^{1/2}\mathbf{x} \Rightarrow \underbrace{\mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2}}_{\mathcal{L}}\mathbf{y} = \lambda\mathbf{y}$. We call the matrix $\mathcal{L} = \mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2}$ the **normalized Laplacian** of graph $G$.

While the vector $\mathbf{x}$ is analogous to temperature and the vector $\mathbf{D}\mathbf{x}$ is analogous to heat from a physical standpoint, the vector $\mathbf{y} = \mathbf{D}^{1/2}\mathbf{x}$ does not have an intuitive physical counterpart. For that reason we avoid using it as it does not allow us to continue having a natural understanding of the problem we are examining.

### 1.3 Completing the convergence

As we saw above, the time derivative of the logarithm of the distance we chose can be bounded by a Rayleigh quotient that corresponds to a generalized eigenvector. By integrating over the time interval $[0, t]$ we can bound the difference of the logarithms between those two points.

$$\begin{aligned}
\frac{d}{dt} \ln\left(d\left(\mathbf{x}(t), \bar{x}\vec{\mathbb{1}}\right)\right) &\leq -2 \min_{\mathbf{x} \perp \vec{\mathbb{1}}} \frac{\mathbf{x}^T \mathbf{L} \mathbf{x}}{\mathbf{x}^T \mathbf{D} \mathbf{x}} \\
&= -2\lambda_2(\mathbf{L}, \mathbf{D}) \Rightarrow \\
d\left(\mathbf{x}(t), \bar{x}\vec{\mathbb{1}}\right) - d\left(\mathbf{x}(0), \bar{x}\vec{\mathbb{1}}\right) &\leq -2\lambda_2(\mathbf{L}, \mathbf{D}) t \Rightarrow \\
\frac{d\left(\mathbf{x}(t), \bar{x}\vec{\mathbb{1}}\right)}{d\left(\mathbf{x}(0), \bar{x}\vec{\mathbb{1}}\right)} &\leq e^{-2\lambda_2(\mathbf{L}, \mathbf{D}) t} \Rightarrow \\
d\left(\mathbf{x}(t), \bar{x}\vec{\mathbb{1}}\right) &\leq e^{-2\lambda_2(\mathbf{L}, \mathbf{D}) t} d\left(\mathbf{x}(0), \bar{x}\vec{\mathbb{1}}\right)
\end{aligned}$$

Notice that the above bound is ineffective when $\lambda_2(\mathbf{L}, \mathbf{D}) = 0$, meaning when the graph is disconnected. Indeed, in that case the distance from the uniform vector might never go below a certain threshold, or even shrink at all.

## 2 Heat kernel via Matrix Exponentiation

In the previous section we obtained a bound on the distance from the uniform vector. In this section, our goal is to provide closed form solutions to the Heat Equation through Matrix Exponentiation.

In Ordinary Differential Equations, for a single variable following the heat equation it's proven that

$$\frac{dx}{dt} = -c \cdot x \Leftrightarrow x = e^{-c \cdot t} + \alpha$$

**Idea:** By defining the equivalent of a matrix exponent we could offer the same kind of closed form solutions to the Heat Equation on graphs.

**Definition 1.** For a real, symmetric matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ we can write

$$e^{\mathbf{A}} = \mathbf{I} + \frac{\mathbf{A}}{1!} + \frac{\mathbf{A}^2}{2!} + \cdots = \sum_{i=0}^{\infty} \frac{\mathbf{A}^i}{i!} \tag{1}$$

We can replace with the eigendecomposition $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$, which satisfies $\mathbf{V}^T\mathbf{V} = \mathbf{I}$. But notice that

$$\mathbf{A}^2 = (\mathbf{V}\mathbf{\Lambda}\mathbf{V}^T)(\mathbf{V}\mathbf{\Lambda}\mathbf{V}^T) = \mathbf{V}\mathbf{\Lambda}\mathbf{I}\mathbf{\Lambda}\mathbf{V} = \mathbf{V}\mathbf{\Lambda}^2\mathbf{V}$$

This property extends to all integer powers of $\mathbf{A}$[2], so we can replace them in equation (1)

[2] It also extends to real powers, but that requires a slightly different proof

$$e^{\mathbf{A}} = \sum_{i=0}^{\infty} \frac{\mathbf{V}\mathbf{\Lambda}^i\mathbf{V}^T}{i!} = \mathbf{V}\sum_{i=0}^{\infty}\left(\frac{\mathbf{\Lambda}^i}{i!}\right)\mathbf{V}^T = \mathbf{V}e^{\mathbf{\Lambda}}\mathbf{V}^T \tag{2}$$

Similarly, we can show that $\frac{d}{dt}e^{\mathbf{A}t} = \mathbf{A}e^{\mathbf{A}t}$. Putting it all back together and replacing $\mathbf{y}(t) = \mathbf{D}^{1/2}\mathbf{x}(t)$ in the Heat Equation it yields

$$\begin{aligned}
\frac{d}{dt}\left(\mathbf{D}^{-1/2}\mathbf{y}(t)\right) &= -\mathbf{D}^{-1}\mathbf{L}\left(\mathbf{D}^{-1/2}\mathbf{y}(t)\right) \Rightarrow \\
\frac{d}{dt}\mathbf{y}(t) &= -\left(\mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2}\right)\mathbf{y}(t) \\
&= -\mathcal{L}\mathbf{y}(t) \Rightarrow \\
\mathbf{y}(t) &= e^{-\mathcal{L}t}\mathbf{y}_0 \Rightarrow \\
\mathbf{D}^{1/2}\mathbf{x}(t) &= \sum_{i=1}^{n} e^{-t\lambda_i(G)}\mathbf{v}_i\mathbf{v}_i^T\mathbf{D}^{1/2}\mathbf{x}_0
\end{aligned}$$

The last line shows that by analyzing $\mathbf{x}_0$ to the eigenvector decomposition, each component shrinks exponentially with a rate equal to the corresponding eigenvalue.