Some things about cutoffs

Ethan Lake

Markov chains which mix quickly, viz. those which have mixing times scaling as the logarithm of the number of states, often approach their steady states in a fashion that displays a sharp cutoff in the 1-norm distance of the time-dependent distribution to the fixed point. Below we will see how this plays out in rather trivial examples involving independent exponentially-relaxing spins and Clifford RU circuits. More sophisticated fast-mixing examples, such as the Ising model at $T > T_c$ and random walks on expander graphs,[1] should behave similarly. These calculations were inspired by discussions with Matteo Ippoliti and Xiaozhou Feng.

Independently relaxing spins

Consider a collection of L independent spins $s_i = \pm 1$, each of which have the distribution

$$p_m(s) = \frac{1+sm}{2} \tag{1}$$

for a time-dependent magnetization[2]

$$m(t) = e^{-\lambda t}. \tag{2}$$

The full distribution at time t is then $p_{m(t)}^{\otimes L}$, and we are interested in the TVD of this distrubtion from the uniform steady state that one obtains as $t \to \infty$. We will bound the TVD using (using quantum notation because why not)

$$1 - F(\rho,\sigma) \leq T(\rho,\sigma) \leq \sqrt{1 - F^2(\rho,\sigma)}, \tag{3}$$

where $T(\rho,\sigma) = \frac{1}{2}||\rho - \sigma||_1$ is the trace norm and $F(\rho,\sigma) = \mathrm{Tr}[\sqrt{\sqrt{\sigma}\rho\sqrt{\sigma}}]$ is the fidelity. Phrasing things in terms of F is useful since F is multiplicative wrt tensor products; for the present case of interest where $\sigma = \mathbf{1}/2^L$ and $\rho = p_{m(t)}^{\otimes L}$, we have

$$F(p_{m(t)}^{\otimes L}, \mathbf{1}/2^L) = \left(\frac{\sqrt{1+m(t)} + \sqrt{1-m(t)}}{2}\right)^L. \tag{4}$$

[1] RWs on n-vertex expanders have $O(1)$ relaxation times and $O(\log n)$ mixing times; this separation between relaxation and mixing is believed to imply a cutoff in many classes of examples (SRWs on bounded-degree graphs in particular).

[2] A different approach to the same problem is to things onto the Ehrnfest urn model (each urn being the spins of a given magnetization), which is well-known to exhibit a cutoff. The approach below however gives us the opportunity to deploy slightly different tools which may be of greater utility when thinking about quantum versions of this problem.

We can then solve for t in terms of F; defining the function

$$\mathsf{T}(F) \equiv \frac{1}{\lambda} \ln \frac{1}{2F^{1/L}\sqrt{1-F^{2/L}}}, \tag{5}$$

one has $\mathsf{T}(F(p_{m(t)}^{\otimes L}, \mathbf{1}/2^L)) = t$.

The mixing time $t_{\mathrm{mix}}(\varepsilon)$ for a given initial distribution[3] $p_{m(0)}^{\otimes L}$ is defined by the first time t such that $T(p_m(t)^{\otimes L}, \mathbf{1}/2^L)$ drops below some fixed constant ε. Using the above bounds for the trace distance, we have

$$\mathsf{T}(1-\varepsilon) \le t_{\mathrm{mix}}(\varepsilon) \le \mathsf{T}(\sqrt{1-\varepsilon^2}). \tag{6}$$

Using the definition of the T function and expanding to leading order in L, some unilluminating algebra gives

$$\frac{1}{2} \ln \frac{L}{2|\ln(1-\varepsilon)|} - \ln 2 \le \lambda t_{\mathrm{mix}}(\varepsilon) \le \frac{1}{2} \ln \frac{L}{|\ln(1-\varepsilon^2)|} - \ln 2, \tag{7}$$

where we have dropped terms going as $O(1/L)$. These bounds give $t_{\mathrm{mix}}(\varepsilon) = \Theta(\ln L)$, viz. $\frac{1}{2}\ln L + c_1 \le t_{\mathrm{mix}}(\varepsilon) \le \frac{1}{2}\ln L + c_2$, where $c_{1,2}$ are ε-dependent $O(1)$ constants.

Consider then what happens at times $t_\alpha = \frac{\alpha}{2\lambda} \ln L$. At these times the fidelity in the $L \to \infty$ limit is

$$\begin{aligned} F(t_\alpha) &= \exp\left(L \ln \frac{\sqrt{1+L^{-\alpha/2}} + \sqrt{1-L^{-\alpha/2}}}{2} \right) \\ &\to \exp(-L^{1-\alpha}/8). \end{aligned} \tag{8}$$

Thus the fidelity has a cutoff at t_1: for any $\delta > 0$ $F(t_{1+\delta})$ is exponentially close to unity, while $F(t_{1-\delta})$ is exponentially close to zero; this in turn implies that $T(p_{m(t)}^{\otimes L}, \mathbf{1}/2^L)$ undergoes a cutoff at t_1 from being exponentially close to 1 to being exponentially close to zero. Furthermore, this cutoff is sharp, having a window whose width is exponentially smaller than t_1. Indeed, let $t_F(\eta)$ denote the time at which the fidelity equals η. Some algebra yields

$$t_F(\eta) = t_1 - \frac{1}{2\lambda} \ln(8 \ln 1/\eta). \tag{9}$$

Therefore for fixed small $0 < \delta < 1$,

$$t_F(1-\delta) - t_F(\delta) = \ln\left(\frac{\ln \delta}{\ln(1-\delta)}\right) = O(1), \tag{10}$$

which implies a similar window on the cutoff of the trace distance. If we define the relaxation period by the window over which F varies from $1/\mathrm{poly}(L)$ to $1-1/\mathrm{poly}(L)$, then from the above the size of the window is $O(\log\log L)$, which as claimed is exponentially smaller than t_1.

It is perhaps worth emphasizing the quasi-nontrivial fact that the cutoff occurs at time $\sim \ln L$, rather than at a time $\sim L^0$: even though the spins are completely decoupled, memory of their initial state survives for an extensively long time (physically, this is due to rare region effects). Therefore we cannot call a system which remembers its intitial conditions for extensively long times a "memory": the times have to grow at least super-logrithmically with L.

[3] Usually one maximizes over initial distributions but here retaining this information will be useful.

Clifford RU circuits

Consider 1D qubit Clifford RU dynamics starting from a product state $|0\rangle^{L_{\rm tot}}$, and examine the RDM $\rho_A(t)$ on a contiguous subsystem of size $L \ll L_{\rm tot}$. We will be interested in how $\rho_A(t)$ approaches its late-time value of $\mathbf{1}/2^L$ as quantified by the trace distance $T(\rho_A, \mathbf{1}/2^L)$. Recall that the RDM for Clifford evolution is

$$\rho_A = \frac{1}{|\mathcal{S}_A|}\Pi_{\mathcal{S}_A}, \tag{11}$$

where $\mathcal{S}_A$ is the logical space of the stabilizer group G_A (viz. the simultaneous $+1$ eigenspace of all elements in G_A), $\Pi_{\mathcal{S}_A}$ is the projector onto $\mathcal{S}_A$, and G_A is the stabilizer group associated with $\rho_A(t)$, viz. the set of stabilizers $S_i = U^\dagger(t) Z_i U(t)$, $i \in A$ such that $(S_i)|_{A^c} = \mathbf{1}_{A^c}$. We then have

$$T(\rho_A, \mathbf{1}/2^L) = \frac{1}{2}\left(\frac{1}{|\mathcal{S}_A|} - \frac{1}{2^L}\right)|\mathcal{S}_A| + \frac{1}{2}2^{-L}(2^L - |\mathcal{S}_A|) = 1 - |\mathcal{S}_A|/2^L = 1 - 2^{S_A - L}, \tag{12}$$

where

$$S_A = -\log \mathrm{Tr}\rho_A^2 \tag{13}$$

is the entropy (using Renyi-2 for concreteness, but the fact that ρ_A is proportional to a projector means that all Renyi entropies are the same). To connect this to the Bures distance strategy above, we note that the fidelity is

$$F(\rho_A, \mathbf{1}/2^L) = 2^{\frac{1}{2}(2\log \mathrm{Tr}\sqrt{\rho_A} - L)} = 2^{(S_A - L)/2}, \tag{14}$$

since $2\log \mathrm{Tr}\sqrt{\rho_A}$ is the $\alpha = 1/2$ Renyi entropy. Therefore in this case the trace distance is $T = 1 - F^2$, which is strictly inbetween the lower bound of $1 - F$ and the upper bound of $\sqrt{1 - F^2}$.

We then need to know how $S_A(t) - L$ evolves under the RU evolution. A decent proxy for this is to use the fact that operator string endpoints undergo biased random walks to write

$$L - S_A(t) \to \frac{L}{2}\left(1 - \mathrm{erf}\left(\frac{v_E t - L}{\sqrt{Dt}}\right)\right), \tag{15}$$

where v_E is the entanglement velocity and D is some $O(1)$ diffusion coefficient. Thus $T(\rho_A, \mathbf{1}/2^L)$ goes from $1 - O(\exp(-L))$ to $O(\exp(-L))$ near time $t_* = L/v_E$, with the window width being parametrically (in L) smaller than t_*.