Hydrodynamic limit of move-to-front rules and LRU caching

Tetsuya Hattori (Keio Univ.)
Kumiko Hattori (Tokyo Metropolitan Univ.)

We study a hydrodynamic limit approach to move-to-front rules. Namely, we consider a stochastic process of particles aligned in a line, each of which jumps randomly to the top, and study a large particle number limit. A scaling limit of the joint empirical distribution of jump rate and position satisfies a system of Burgers type partial differential equations. Results are applied to the search cost probabilities for the least-recently-used caching in the data theory of computer sciences.

1. Move-to-front rules. Let N be a positive integer, and $\mathcal{S}_N$ a set of all permutations of $1, 2, \cdots, N$, and define a Markov process

$$X^{(N)}(t) = (X_1^{(N)}(t), \cdots, X_N^{(N)}(t)), \quad t \geqq 0,$$

with the state space $\mathcal{S}_N$, as follows. For each $i = 1, 2, \ldots, N$, let $\tau_{i,j}^{(N)}$, $j = 1, 2, \cdots$, be an increasing sequence of random variables (jump times, to be specified shortly), and we define $X^{(N)}(t)$ to be constant in t for $t \notin \{\tau_{i,j}^{(N)} \mid i = 1, 2, \cdots, N,\ j = 1, 2, \cdots\}$. At a jump time $t = \tau_{i,j}$, we define $X_i^{(N)}(\tau_{i,j}) = 1$, and for $i' \neq i$,

$$X_{i'}^{(N)}(\tau_{i,j}) = X_{i'}^{(N)}(\tau_{i,j} - 0) + \begin{cases} 1, & \text{if } X_{i'}^{(N)}(\tau_{i,j} - 0) < X_i^{(N)}(\tau_{i,j} - 0), \\ 0, & \text{if } X_{i'}^{(N)}(\tau_{i,j} - 0) > X_i^{(N)}(\tau_{i,j} - 0). \end{cases}$$

For convenience, we put $\tau_{i,0}^{(N)} = 0$ $(\forall i)$, and we define $\{\tau_{i,j+1}^{(N)} - \tau_{i,j}^{(N)},\ j = 0, 1, 2, \cdots\}$ to be independent in i and j, identical in distribution for all j, whose distribution is the exponential distribution with parameter $w_i^{(N)} > 0$: $\mathrm{P}[\, \tau_{i,1}^{(N)} > t \,] = \exp(-w_i^{(N)} t)$. Note that as in the standard Poisson process, with probability 1 the jump times are different for different (i, j). This completes the definition of the process $X^{(N)}$.

In the following, we regard $X^{(N)}$ as an N particle system aligned on a single line, with the suffix i in $X_i^{(N)}(t)$ standing for the label of the particle and $X_i^{(N)}(t)$ denoting the position (rank) of the particle i at time t.

2. Hydrodynamic limit. We embed $\mathcal{S}_N \subset \mathbb{R}_+$ and scale by N to consider a particle system in an interval $[0, 1)$; $Y_i^{(N)}(t) := \dfrac{1}{N}(X_i^{(N)}(t) - 1)$. Note that $y_C^{(N)}(t) = \dfrac{1}{N}\sharp\{i \mid \tau_{i,1} \leqq t\} \in [0, 1)$ is the boundary of particles which jumped to the top position and those which has not jumped up to time t. In the following we denote by δ_a the unit distribution concentrated on a, and assume $\lambda^{(N)} := \dfrac{1}{N}\displaystyle\sum_{i=1}^{N} \delta_{w_i^{(N)}} \to \lambda$ weakly as $N \to \infty$, for a probability distribution λ.

Proposition 1([1]). $y_C^{(N)}(t) \to y_C(t) := 1 - \displaystyle\int_0^\infty e^{-wt}\, \lambda(dw)$ $(N \to \infty$, in prob.). $\diamond$

Consider a joint empirical distribution $\mu_t^{(N)} = \dfrac{1}{N}\displaystyle\sum_{i=1}^{N} \delta_{(w_i^{(N)}, Y_i^{(N)}(t))}$.

Theorem 2([1]). Assume $\displaystyle\int_0^\infty w\lambda(dw) < \infty$ and $\lambda(\{0\}) = 0$, and assume that the initial distribution $\mu_0^{(N)}$ determined by the initial configuration $Y^{(N)}(0) = y^{(N)}$ converges weakly to a distribution μ_0 as $N \to \infty$. Then for each $t > 0$, there exists a deterministic distribution μ_t such that $\mu_t^{(N)} \to \mu_t$ as $N \to \infty$. μ_t is given by

$$U(dw, y, t) := \mu_t(dw, [y, 1)) = \begin{cases} \lambda(dw)\, e^{-wt_0(y)}, & y < y_C(t), \\ U(dw, \hat{y}(y, t), 0)\, e^{-wt}, & y > y_C(t), \end{cases}$$

where, $t = t_0(y)$ is the inverse function of $y = y_C(t)$, and $\hat{y}(y,t)$ is the inverse function in y of $y_C(y,t) = 1 - \int_y^1 \int_0^\infty e^{-wt} \mu_0(dw, dz)$. $\diamond$

The assumption $\int_0^\infty w\lambda(dw) < \infty$ is unnecessary for the convergence at $y > 0$.

3. Burgers type equation. We succeeded in proving **Theorem 2** by guessing the explicit formula for μ_t correctly, and then by proving the convergence. The explicit formula hence is of importance, which we found as a solution to a following system of PDEs. Consider the case where there are at most countable types of jump rates; $\lambda = \sum_\alpha \rho_\alpha \delta_{f_\alpha}$, where f_α and ρ_α are positive constants, satisfying $\sum_\alpha \rho_\alpha = 1$.

Proposition 3([2]). $U_\alpha(y,t) := U(\{f_\alpha\}, y, t) = \mu_t(\{f_\alpha\}, [y,1))$ is a unique classical time global solution to the following initial value problem: $\frac{\partial U_\alpha}{\partial t}(y,t) + \sum_\beta f_\beta U_\beta(y,t) \frac{\partial U_\alpha}{\partial y}(y,t) = -f_\alpha U_\alpha(y,t)$, $(y,t) \in [0,1) \times [0,\infty)$, $\alpha = 1, 2, \cdots$, with boundary conditions $U_\alpha(0,t) = \rho_\alpha$, $t \geqq 0$, $\alpha = 1, 2, \cdots$. For each α, the initial data $U_\alpha(y,0) = U_\alpha(y)$, $0 \leqq y < 1$, are smooth, non-negative, non-decreasing, satisfying $\sum_\beta f_\beta U_\beta(0) < \infty$ and $\sum_\beta U_\beta(y) = 1 - y$. $\diamond$

This system is solved by a standard method of characteristic curves, with explicit formula containing inverse functions such as t_0 of the characteristic curve y_C. Hence the idea of hydrodynamic limit is of relevance to the results. (Incidentally, the method of characteristic curves gives time local solutions, while the assumptions on initial data satisfies no-shock wave condition, implying time global solution.)

4. Search cost. There is a large number of studies concerning move-to-front rule in the context of data theory in computer sciences. The LRU (least-recently-used) caching as a data allocation algorithm in computer memory or web page browsing is equivalent to move-to-front rule, with a data request corresponding to a particle jumping to the top position. The search cost C_N defined as the position of the first requested data just before the request is of interest. See [4] for references. We have $\frac{1}{N} C_N = Y^{(N)}_{Q^{(N)}(t)}(t)$, where $Q^{(N)}(t)$ is the label of the particle which jumped first after time t, to which we can apply **Theorem 2** to obtain, for example, $\lim_{N\to\infty} \mathrm{P}_t[\ \frac{1}{N} C_N(t) > x\ ] = \frac{\iint w\,\mu_{y,t}(dw)\,dy}{\int w\,\lambda(dw)}$.

Stationary distribution $\mathrm{E}_\infty[\ \cdot\ ]$ for $N < \infty$ has naturally been studied since the model's first appearance in the literature [5]. By having the initial configuration $y^{(N)}$ distribute under $\mathrm{E}_\infty[\ \cdot\ ]$, we can handle the stationary distribution $\mu_\infty^{(N)} = \mathrm{E}_\infty[\ \mu_0^{(N)}\ ]$ for the joint jump rate and position distribution in our framework. **Theorem 2** then implies, for example, $\lim_{N\to\infty} \mathrm{P}_\infty[\ \frac{1}{N} C_N > x\ ] = \frac{\int w e^{-w t_0(x)} \lambda(dw)}{\int w \lambda(dw)}$.

References.

[1] K. Hattori, T. Hattori, Stochastic Processes & Applications **119** (2009) 966–979.
[2] K. Hattori, T. Hattori, Funkcialaj Ekvacioj (2009), to appear.
[3] K. Hattori, T. Hattori, preprint (2008).
[4] K. Hattori, T. Hattori, preprint (2009).
[1–4] available at `http://web.econ.keio.ac.jp/staff/hattori/liamazn.htm`.
[5] M. L. Tsetlin, Russian Math. Surv. **18**(4) (1963) 1–27.