数理解析研究所講究録
第 1952 巻 2015 年 86-91

Phase transitions for controlled Markov chains on infinite graphs*

Naoyuki Ichihara
Department of Physics and Mathematics
Aoyama Gakuin University†

Abstract

The objective of this note is to give a survey of our recent paper [6]. We are concerned with certain phase transition phenomena arising in some parametrized maximization problems for controlled Markov chains on infinite graphs. We prove that, there exists a critical value of the parameter such that the recurrence/transience of the optimal trajectory changes around this critical point.

1 Main results

Let S be a countably infinite set with a fixed reference point $x_0 \in S$ (e.g., $S=\mathbf{Z}^{d}$ with $x_0=0$), and let $p=p(x,y)$ be a given stochastic matrix on S. Set $B:=\{(x,y)\in S\times S\,|\,p(x,y)>0\}$. Then the pair (S,B) forms an infinite graph equipped with the graph distance $d(x,y):=\inf\{n\geq 0\,|\,p^{n}(x,y)>0\}$, where p^{n} denotes the n-product of the stochastic matrix p. Notice here that $d(x,y)\neq d(y,x)$, in general. Throughout this note, we impose the following conditions:

(A1) (a) $p=p(x,y)$ is irreducible on S, i.e., the graph (S,B) is strongly connected.

(b) There exists an $M>0$ such that $d(y,x)\leq M$ for all $(x,y)\in B$.

*This research was supported in part by JSPS KAKENHI Grant No. 24740089.
†5-10-1 Fuchinobe, Chuo-ku, Sagamihara-shi, Kanagawa 252-5258, Japan.

(c) There exists an $\varepsilon_0 > 0$ such that $p(x, y) \geq \varepsilon_0$ for all $(x, y) \in B$.

(d) $\frac{1}{n}(p^n d)(x_0) \to 0$ as $n \to \infty$, where $d(x) := d(x_0, x)$.

Let Q be the totality of stochastic matrices $q = q(x, y)$ on S such that $q(x, y) = 0$ for all $(x, y) \notin B$. For each $q \in Q$, we denote by $((X_n)_{n \geq 0}, (\mathbf{P}_x^q)_{x \in S})$ the associated discrete-time, time-homogeneous Markov chain on S, which we shall call "q-chain" for short. Regarding q as a stationary control policy, we consider the following maximization problem with real parameter $\beta \geq 0$:

$$\begin{aligned} &\text{Maximize} \quad J(q; \beta) := \limsup_{T \to \infty} \frac{1}{T} \mathbf{E}_{x_0}^q \left[\sum_{n=0}^{T-1} (\beta r(X_n) - c(X_n, q)) \right], \\ &\text{subject to} \quad q \in Q, \end{aligned} \tag{1}$$

where $r : S \times \mathbf{R}$ is a reward function, while $c : S \times Q \to \mathbf{R}$ stands for a penalty function defined by

$$c(x, q) = \frac{1}{a(x)} \sum_{y \in S} q(x, y) \log \frac{q(x, y)}{p(x, y)}, \quad x \in S, \tag{2}$$

for some $a = a(x) : S \to \mathbf{R}$. Hereafter, we always assume the following:

(A2) There exist finite $x_1, \ldots, x_l \in S$ and $\alpha_1, \ldots, \alpha_l > 0$ such that $r(x) = \sum_{i=1}^{l} \alpha_i \delta_{x_i}(x)$ for all $x \in S$.

(A3) There exist $\kappa_1, \kappa_2 > 0$ such that $\kappa_1 \leq a(x) \leq \kappa_2$ in S.

By definition, c is nonnegative in $S \times Q$ and strictly convex with respect to q. It is also not difficult to verify that $c(x, q) = 0$ if and only if $q = p$.

Our interest is to investigate qualitative properties, with respect to β, of the optimal value

$$\Lambda(\beta) := \sup_{q \in Q} J(q; \beta), \tag{3}$$

as well as the corresponding optimal control policy. Since the value $\Lambda(\beta)$ is determined as a trade-off between reward $r(x)$ and penalty $c(x, q)$, it is natural to expect that certain phase transitions occur at some β. It turns out that, if the p-chain is transient, then there exists a $\beta_c > 0$ such that $\Lambda(\beta) = 0$ for $\beta \in [0, \beta_c]$ and $\Lambda(\beta) > 0$ for $\beta \in (\beta_c, \infty)$. Loosely speaking, if β is small, then our optimal strategy is to choose p at any stage (and obtain $\Lambda(\beta) = J(p; \beta) = 0$) since the reward does not compensate the penalty incurred by $q \neq p$. In contrast, if β is large, then a suitable policy $\bar{q} \neq p$ allows one to obtain some positive return $\Lambda(\beta) = J(\bar{q}; \beta) > J(p; \beta) = 0$. Hence, there exists a threshold $\beta = \beta_c$ at which the controller switches his/her strategy from p to another $\bar{q} \neq p$, and several qualitative properties concerning (1) change in the vicinity of $\beta = \beta_c$.

In [6], we justify such phase transition phenomena in the framework of controlled Markov chains, or Markov decision processes, on S. To this end, we introduce the optimality equation associated with (1). For each $f : S \to \mathbf{R}$, we define the function $H[f] : S \to \mathbf{R}$ by

$$H[f](x) := \sup_{q \in Q}\{(qf)(x) - c(x,q)\}, \quad x \in S, \tag{4}$$

and consider the following difference equation:

$$\lambda + W(x) = H[W](x) + \beta r(x) \quad \text{in } S, \quad W(x_0) = 0, \tag{5}$$

where the unknown is the pair of a real constant λ and a function $W = W(x)$ on S. Note that we impose the constraint $W(x_0) = 0$ to avoid the ambiguity of additive constants with respect to W. We next set

$$\lambda^*(\beta) := \inf\{\lambda \in \mathbf{R} \,|\, \text{There exists a supersolution } W \text{ of } (5)\}. \tag{6}$$

Recall that a function $W : S \to \mathbf{R}$, or a pair (λ, W), is called a supersolution (resp. subsolution) of (5) if

$$\lambda + W(x) \geq H[W](x) + \beta r(x) \ \text{ in } S, \quad W(x_0) \geq 0 \qquad (\text{resp. } \leq). \tag{7}$$

As usual, W is said to be a solution of (5) if it is both sub- and supersolutions.

We are now in position to state our main results. We first give the solvability of (5).

Theorem 1.1 (see Theorem 2.1 of [6]). *For any $\beta \geq 0$, there exists a solution W of (5) with $\lambda = \lambda^*(\beta)$ such that $\sup_S W < \infty$. Moreover, there exists a $\beta_c \geq 0$ such that $\lambda^*(\beta) = 0$ for $\beta \in [0, \beta_c]$ and $\lambda^*(\beta) > 0$ for $\beta \in (\beta_c, \infty)$.*

Our second result is concerned with the characterization of $\Lambda = \Lambda(\beta)$ in terms of the constant $\lambda^*(\beta)$ in (6). Let $\bar{q} = \bar{q}(x,y)$ be the stochastic matrix on S that maximizes the right-hand side of (4) for all $x \in S$, i.e., the one which satisfies

$$H[f](x) = (\bar{q}f)(x) - c(x,\bar{q}), \quad x \in S. \tag{8}$$

It is known (e.g., Lemma 3.3 of [6]) that such $\bar{q}$ exists uniquely, and that it is irreducible on S. Hereafter, we use the notation $\bar{q} = \operatorname{argmax} H[f]$ to emphasize the dependence of $\bar{q}$ on f. Then, our second main result can be stated as follows.

Theorem 1.2 (c.f., Theorem 2.2 of [6]). *Let $\lambda^* = \lambda^*(\beta)$ and $\Lambda = \Lambda(\beta)$ be the constants defined by (6) and (3), respectively. Then $\lambda^*(\beta) = \Lambda(\beta)$ for all $\beta \geq 0$. Moreover, Let W be the solution of (5) given in Theorem 1.1, and let $\bar{q} = \operatorname{argmax} H[W]$. Then $\Lambda(\beta) = J(\bar{q}; \beta)$.*

We now turn to our third result concerning phase transition phenomena arising in (1). More precisely, let β_c be the constant given in Theorem 1.1. Then the positivity of β_c is equivalent to the transience of the p-chain. One can also see that the recurrence and transience of the Markov chain associated with the optimal control policy $\bar{q} = \operatorname{argmax} H[W]$ changes whether $\beta > \beta_c$ or $\beta < \beta_c$.

Theorem 1.3 (c.f., Theorem 2.3 of [6]). *Let $\beta_c \geq 0$ be the constant given in Theorem 1.1. Then, $\beta_c > 0$ if and only if the p-chain is transient. Moreover, for any solution W of (5) with $\lambda = \lambda^*(\beta)$, the Markov chain associated with $\bar{q} = \operatorname{argmax} H[W]$ is transient for $\beta < \beta_c$ and positive recurrent for $\beta > \beta_c$.*

By virtue of Theorems 1.1, 1.2, and 1.3, we observe that the stationary policy p is optimal for $\beta \leq \beta_c$, while it is not optimal for $\beta > \beta_c$. This agrees with our intuition.

2 Some additional remarks

We first mention that, if $a = a(x)$ in (2) is constant in S, say $a(x) \equiv 1$, then the optimality equation (5) can be transformed into a linear equation. More specifically, suppose that $a(x) \equiv 1$, namely, $c(x, q)$ is equal to the so-called relative entropy function with respect to p:

$$c(x,q) = \sum_{y \in S} q(x,y) \log \frac{q(x,y)}{p(x,y)}, \quad x \in S.$$

Then, one can see that W is a solution of (5) if and only if $\phi(x) = e^{W(x)}$ is a positive solution to the linear stationary equation

$$\sum_{y \in S} e^{\beta r(x)} p(x,y) \phi(y) = e^{\lambda} \phi(x), \quad x \in S. \tag{9}$$

In this sense, the optimality equation (5) can be regarded as a nonlinear extension of the linear eigenvalue problem (9). Furthermore, by regarding (9) as the limit of the recursive equation

$$\phi(T,x) = \sum_{y\in S} e^{\beta r(x)} p(x,y)\phi(T-1,y) \quad \text{in } \mathbf{N}\times S, \quad \phi(0,x)\equiv 1, \tag{10}$$

and taking into account the large deviation theory, one can expect that

$$\begin{aligned}\lim_{T\to\infty} \frac{\log\phi(T,x)}{T} &= \lim_{T\to\infty} \frac{1}{T}\log \mathbf{E}_x^p \left[\exp\left(\beta\sum_{n=0}^{T-1} r(X_n)\right)\right] \\ &= \inf\{\lambda\in\mathbf{R} \mid (9) \text{ has a positive solution } \phi\}.\end{aligned} \tag{11}$$

The relations (11) can be justified as a corollary of Theorem 1.2. We emphasize, however, that the transformation above does not work when $a(x)$ in (2) is not constant, so that our maximization problem is essentially nonlinear.

There is an extensive literature devoted to limit theorems of type (11). In the case where $S = \mathbf{Z}^d$, the paper [1] studies the continuous-time counterpart of (9)-(11) in connection with discrete homopolymer models in statistical physics. More precisely, suppose that $S = \mathbf{Z}^d$ and $r(x) = \delta_0(x)$, where $\delta_y : S \to \mathbf{R}$ is defined by $\delta_y(x) = 1$ for $x = y$ and $\delta_y(x) = 0$, otherwise. Let $X = (X_t)_{t\geq 0}$ be the continuous-time random walk on $\mathbf{Z}^d$ generated by the discrete Laplacian Δ. Furthermore, let $F(\beta)$ be the "free energy" of X defined by

$$F(\beta) := \lim_{T\to\infty} \frac{1}{T}\log \mathbf{E}_0 \left[\exp\left(\beta\int_0^T \delta_0(X_t)dt\right)\right].$$

Then, it is proved in [1] that $F(\beta)$ is identical with the spectral function $\sup\sigma(L)$ of the discrete Scrödinger operator $L := \Delta + \beta\delta_0$ on $l^2(\mathbf{Z}^d)$, that there exists a $\beta_c \geq 0$ such that $F(\beta) = 0$ for $\beta\in[0,\beta_c]$ and $F(\beta) > 0$ for $\beta\in(\beta_c,\infty)$, and that $\beta_c = 0$ for $d = 1,2$ and $\beta_c > 0$ for $d \geq 3$ (see also [3] for the spectral analysis of L). In [6], we rediscover these facts as a particular case of our main results, although we only deal with the discrete-time case. We point out here that, if $S = \mathbf{Z}^d$ and $p = p(x,y)$ generates a symmetric simple random walk on $\mathbf{Z}^d$, then (9) is nothing but a discrete-time variant of the stationary Schrödinger equation $L\phi = \lambda\phi$ in $\mathbf{Z}^d$.

As the final remark, we mention that [6] can be regarded as a discrete version of [4, 5], where the same type of phase transitions are discussed for continuous-time controlled diffusions in $\mathbf{R}^d$ (see also [2] for a continuous version of [1]). Our interest in [6] is to establish an effective method peculiar in the discrete case. Apart from technical details, we also obtain some results that have not beed studied in the continuous model (see Theorem 6.6 in [6]). Typically, we are able to estimate the value of β_c, and this seems to be new to the best of our knowledge.

References

[1] M. Cranston, S. Molchanov, On phase transitions and limit theorems for homopolymers, Probability and mathematical physics, CRM Proc. Lecture Notes, 42, Amer. Math. Soc., Providence, RI, 2007, pp. 97–112.

[2] M. Cranston, L. Koralov, S. Molchanov, B. Vainberg, Continuous model for homopolymers, J. Funct. Anal., 256 (2009), pp. 2656–2696.

[3] F. Hiroshima, I. Sasaki, T. Shirai, A. Suzuki, Note on the spectrum of discrete Schrödinger operators, J. Math-for-Ind., 4B (2012), pp. 105–108.

[4] N. Ichihara, Criticality of viscous Hamilton-Jacobi equations and stochastic ergodic control, J. Math. Pures Appl., 100 (2013), pp. 368–390.

[5] N. Ichihara, The generalized principal eigenvalue for Hamilton-Jacobi-Bellman equations of ergodic type, to appear in Ann. Inst. H. Poincaré Anal. Non Linéaire, to appear.

[6] N. Ichihara, Phase transitions for controlled Markov chains on infinite graphs, preprint.