The Annals of Applied Probability
1993, Vol. 3, No. 4, 1219–1240

OPTIMALITY OF MOVE-TO-FRONT FOR SELF-ORGANIZING DATA STRUCTURES WITH LOCALITY OF REFERENCES

BY PHILIPPE CHASSAING

Université de Nancy

In papers about self-organizing data structures, it is often mentioned that the assumption of independence of successive requests of keys should be relaxed and that the dependence should assume the form of a locality phenomenon. In this setting, the move-to-front rule is considered to be of interest, but no optimality result concerning this rule has yet appeared. In this paper we assume that the sequence of required keys is a Markov chain with a transition kernel P and we consider the class $\mathscr{F}^*$ of stochastic matrices P such that move-to-front is optimal among on-line rules, with respect to the stationary search cost. We give properties of $\mathscr{F}^*$ that bear out the usual explanation of optimality of move-to-front by a locality phenomenon exhibited by the sequence of required keys. We explicitly produce a large subclass of $\mathscr{F}^*$, while showing that in some cases move-to-front is optimal with respect to the speed of convergence toward stationary search cost.

1. Introduction. Let us describe a simple example of a self-organizing sequential search data structure. Let $S = \{1, 2, \dots, M\}$ be a set of items. Let us also assume that these items are stored in different places and that the set $\mathscr{P}$ of places is $\{1, 2, \dots, M\}$. When an item is required, it is searched for in place 1, then, if not found, in place 2, and so on, and a cost p is incurred if the item is finally found in place p. Once the item has been found, a control is made on the search process by replacing the item in a wisely chosen place, in order to decrease the search cost: Very often, the accessed item is replaced closer to place 1, in such a way that the most frequently accessed items spend most of their time near place 1. For instance, according to the rule *transpose*, the accessed item is replaced one step ahead, while, according to the rule *move-to-front*, it is replaced at the front of the list. When doing this, we must free the new position h of the accessed item by pushing the items remaining between the old position k and the new position h, the nonaccessed items retaining their relative order, as in Figure 1.

Let $F = (F_n)_{n \geq 1}$ be the sequence of required keys. In most studies, the starting assumption is that F is a sequence of i.i.d. random variables, but in general it is asserted in the course of the study that such an assumption is a crude approximation that should be relaxed (see [12], [14]–[18], [22]), and

Received June 1992; revised April 1993.

AMS 1991 *subject classifications*. Primary 68P05, 90C40; secondary 60J10.

Key words and phrases. Controlled Markov chain, Bellman optimality condition, self-organizing data structure, sequential search, locality.

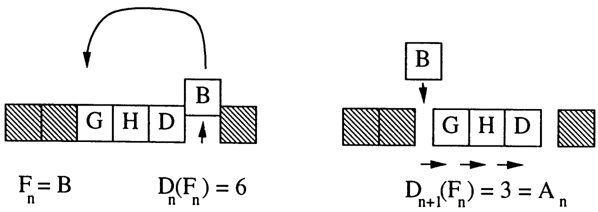

FIG. 1.

some authors (see [12], [14]–[16], [18], [19]) suggest more precisely to consider a Markovian dependence. We shall thus assume that F is a Markov chain, with state space S and transition kernel $P = [p_{i,j}]$. Let D_n be the disposition of keys in $\mathscr{P}$, that is, the one-to-one mapping from S onto $\mathscr{P}$ giving the position of each item after the replacement of F_{n-1}; D_0 is the initial disposition and A_n denotes the control made on the system when F_n has been found, that is, the place where F_n has been replaced.

Because D_{n+1} is completely determined once (D_n, F_n, A_n) is given, the process $X = (X_n)_{n \geq 0}$, where $X_n = (D_n, F_n)$, is a controlled Markov chain with finite state space $S' = \mathscr{S}_M \times S$ and finite control space $A = \mathscr{P}$ (here $\mathscr{S}_M$ is the symmetric group of order $M!$). The cost $c(i, a)$ incurred when the system is in state $i = (x, k)$ and when the control a is chosen, is $x(k)$. Such a controlled Markov chain is generally called a *self-organizing data structure*, and the decision rules (or policies) are called *self-organizing sequential search heuristics* or *replacement rules*.

2. Classification of replacement rules. In the setting of controlled Markov chains, the more general class of rules is the class $\mathscr{C}$ of *admissible* rules (see [8]): A rule R is said to be admissible if the decision A_n is chosen at random in the action space A, but following a distribution $R(H_n)$ depending on the history H_n of the process up to time n, where

$$H_n := (X_0, A_0, X_1, A_1, \ldots, X_{n-1}, A_{n-1}, X_n).$$

Admissible rules are thus, in computer science language, *on-line* rules, in the more general usage of the term. An admissible rule R is thus a mapping from the space of histories to the space of probability measures on A, and the law of the process $(X_n, A_n)_{n \geq 0}$ is well defined once the initial state, let us say i, and the decision rule, let us say R, are given: Probabilities and expectations with respect to this law are denoted by $\mathbb{P}_{R,i}$ and $\mathbb{E}_{R,i}$.

In the special case studied here, the class $\mathscr{C}$ contains every replacement rule mentioned in the literature, as far as I know. The subclass $\mathscr{C}_D$ of deterministic markovian stationary rules plays a special part because, for the usual criteria, if a rule is optimal in the subclass $\mathscr{C}_D$, it is optimal in the class

$\mathscr{C}$ (see [8]): A rule R is in $\mathscr{C}_D$ if there exists a mapping f from S' to A, such that

$$R(H_n) = \delta_{f(X_n)}$$

for each n and each H_n. Then we set $R := (f)$.

P being usually unknown, the real problem is to find a rule optimal for any P, but this seems hopeless, at least when the F_n are dependent (see [6]). Another approach could be as follows: Assume that P is known but that the files have been secretly renumbered, which is an (imperfect) way to simulate ignorance of P. In this way, if the apparent state of the system is (x, k), the state with respect to the old numbering will be $(x \circ \sigma^{-1}, \sigma(k))$ for a wisely chosen permutation σ and the interesting rules are thus the rules of $\mathscr{C}$ that are insensitive to renumbering; that is, the rules that are stable under the action, so defined, of the symmetric group $\mathscr{S}_M$ on S'. The class of such *key ignoring* rules will be denoted by $\mathscr{C}_{\mathrm{KI}}$ (see [4] and [14]). A member of the intersection $\mathscr{C}_L$ of $\mathscr{C}_D$ and $\mathscr{C}_{\mathrm{KI}}$ will be called a *library-type* rule, after the works of Letac [20] and Diès [9], because it has the following characterization (see [4] and [5]):

PROPOSITION 2.1. *A rule (f) of $\mathscr{C}_D$ belongs to $\mathscr{C}_L$ iff $f(x, k)$ depends only on $x(k)$.*

Representative members of $\mathscr{C}_L$ are transpose (R_T), move-to-front (R_F) and move-to-back (R_B). The library-type rules have a low computational cost and require no memory storage. In contrast, the rule *count*, or *counter scheme* (see [2], [3], [10], [13]–[15], [17], [22], [23]), which belongs to $\mathscr{C}_{\mathrm{KI}}$ but not to $\mathscr{C}_D$, is usually discarded because it requires heavy additional storage space. It is, however, quite an interesting candidate because, in the independent case, its stationary search cost is optimal (for each P), being equal to that of the optimal static ordering and, even more surprising, for each n, the average search cost of *count* at step n is optimal in $\mathscr{C}_{\mathrm{KI}}$ (see [14]). In [14] the exact counterpart of the counter scheme in the Markovian case is suggested: At the nth request, if $F_n = k$, replace the files j by decreasing order of $p_{k,j}$, or rather, P being unknown, by decreasing order of the best estimates one can get at that step for $p_{k,j}$, from the sample $(F_0, F_1, \dots, F_n)$. Its stationary search cost is optimal, but, in contrast with the counter scheme for independent references, totally prohibitive auxiliary search costs are incurred, making the rule ineffective (see [6]). In this paper, we thus do not attempt to search for a rule optimal for any P: Many authors (see [2], [17], [22], [23]) suggest that a possible dependence of requests generally assumes the form of a locality phenomenon. According to them, in that case, move-to-front is the first candidate to be considered. We let $\mathscr{F}^*$ be the class of stochastic matrices P such that move-to-front is optimal in $\mathscr{C}$ (and thus a fortiori in $\mathscr{C}_{\mathrm{KI}}$ or in $\mathscr{C}_L$), with respect to the stationary search cost. Thus, this paper aims to give properties of $\mathscr{F}^*$ that bear out the widely accredited explanation of optimality of move-to-front by a locality phenomenon shown by the sequence of required keys, for instance, by exhibiting a large intersection between $\mathscr{F}^*$ and the class of stochastic matrices showing locality patterns as understood in [1].

The optimality results obtained so far for library-type rules once again concern the independent case and differ basically from the Markovian case: They show that transpose has a lower search cost than move-to-front, or that, using a very special law for the F_n, transpose is optimal in a subclass of $\mathscr{C}_L$ (see [9], [15], [21], [22], [24]). In the same spirit, Chung, Hajela and Seymour [7] propose bounds for the ratio ϕ_R/ϕ_{OPT}, under the irm, when R is move-to-front or transpose and OPT is the optimal static ordering.

3. Comparing performances of rules. In order to compare performance of rules, we shall retain two classical criteria: the stationary search cost $\phi_R(i)$,

$$\phi_R(i) = \liminf_n \frac{1}{n} \mathbb{E}_{R,i}(C_1 + \cdots + C_n),$$

where C_n is the cost incurred at step n $[C_n = c(X_n, A_n)]$, and the discounted cost $\Psi_R(i, \beta)$,

$$\Psi_R(i, \beta) = \sum_{n \geq 0} \beta^n \mathbb{E}_{R,i}(C_n), \qquad 0 < \beta < 1$$

(see [8]).

DEFINITION 3.1. R is called *optimal* if for any R' in $\mathscr{C}$, the function $\phi_R(\cdot)$ is less than $\phi'_R(\cdot)$ and 1^- *optimal* if there exists β_0 in $[0, 1[$ such that for any R' in $\mathscr{C}$,

$$\Psi_R(\cdot, \beta) \leq \Psi_{R'}(\cdot, \beta) \quad \forall \beta \in [\beta_0, 1[.$$

Let us consider a rule R in $\mathscr{C}_D$. Under $\mathbb{P}_{R,i}$, $X = (X_n)_{n \geq 0}$ is a Markov chain, $\phi_R(i)$ is a true limit and

$$\Psi_R(\beta) = \frac{1}{1-\beta}\phi_R + v_R + \varepsilon(\beta) \quad \text{with} \quad \varepsilon(\beta) \to 0 \quad \text{when} \quad \beta \to 1^-$$

(see [8], page 69). So 1^- optimality of a rule R of $\mathscr{C}_M$ entails optimality of R. It entails, as well, the optimality of v_R among the rules where the stationary search cost is optimal. Note that v_R is just the *overwork* as defined by Bitner [3]; that is, we have

$$\mathbb{E}_{R,i}(C_1 + \cdots + C_n) = n\phi_R(i) + v_R(i) + \varepsilon(n),$$

when, under R, the ergodic sets of X are not cyclic. The rate of convergence toward ϕ_R is also considered in [11], [16] and [22]. One can also view v_R as a starting cost (cf. [8], pages 66–67). Thus a 1^- optimal rule connects the optimality of the stationary search cost to the starting cost optimality.

Let $y_R(i, a) := (y_R^1(i, a), y_R^2(i, a))$ be defined by

$$y_R^1(i, a) := \sum_{j \in S'} p_{i,j}(a)\phi_R(j) - \phi_R(i),$$

$$y_R^2(i, a) := c(i, a) + \sum_{j \in S'} p_{i,j}(a) v_R(j) - v_R(i) - \phi_R(i).$$

Let A_i (resp. B_i and C_i) be the set of actions a such that $y_R(i, a)$ is lexicographically lower than 0 (resp. 0 and greater than 0), and note that $y_R(i, R(i))$ is zero [cf. (6.3)]. Let us recall some classical criteria for stationary cost optimality and for 1^- optimality:

BELLMAN'S OPTIMALITY CONDITIONS. If A_i is void for each i, R is said to be *Bellman-optimal* and if, further, for each i, $B_i = \{R(i)\}$, R is said to be *strictly Bellman-optimal*. Bellman-optimality entails optimality and strict Bellman-optimality entails 1^- optimality.

PROPOSITION 3.1. *If $R'(i)$ belongs to $A_i \cup B_i$ for each i, then*

$$\phi_{R'}(\cdot) \le \phi_R(\cdot).$$

In the sequel, $\mathscr{F}^*$ (resp. $\mathscr{F}^B$ and $\mathscr{F}^{BS}$) will denote the classes of stochastic $M \times M$ matrices P such that if P is the transition matrix of $F = (F_n)_{n \ge 0}$, then move-to-front is optimal (resp. Bellman-optimal and strictly Bellman-optimal), and $\mathscr{B}^*$ (resp. $\mathscr{B}^B$ and $\mathscr{B}^{BS}$) will denote the classes of stochastic matrices having the same properties with respect to move-to-back.

4. Results.

The results of this paper follow.

THEOREM 1. (i) *Let $P = [p_{m,n}]$ be an element of $\mathscr{F}^*$ and let $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_M)$ be in $[0,1]^M$. Let P_α be the matrix with the general term*

$$p^\alpha_{m,n} = \begin{cases} \alpha_m p_{m,n}, & \text{if } m \ne n, \\ 1 - \alpha_m + \alpha_m p_{m,m}, & \text{if } m = n. \end{cases}$$

Then P_α is in $\mathscr{F}^*$.

(ii) *The result remains valid when replacing $\mathscr{F}^*$ with $\mathscr{B}^*$ and $[0,1]^M$ with $[1, +\infty[^M$ if P_α is still a stochastic matrix [i.e., if $\sup_m \alpha_m(1 - p_{m,m}) \le 1$].*

THEOREM 2. (i) *Let P be an element of $\mathscr{F}^B$ and let $\alpha = (\lambda, \lambda, \dots, \lambda)$ be in $[0, 1[^M$. Then P_α is in $\mathscr{F}^{BS}$.*

(ii) *The result remains the same when replacing $\mathscr{F}$ with $\mathscr{B}$ and $[0,1[^M$ with $]1, +\infty[^M$ [if $\lambda \le \sup_m (1 - p_{m,m})^{-1}$].*

The effect of the perturbations on P, in parts (i) of these theorems, is to increase the locality phenomenon and in parts (ii), to decrease it.

COROLLARY. *The stochastic matrices P with general term*

$$p_{m,n} = \begin{cases} \beta_m, & \text{if } m \ne n, \\ \alpha_m, & \text{if } m = n \end{cases}$$

are in $\mathscr{F}^*$ if $\alpha \ge \beta$ and in $\mathscr{B}^*$ if $\alpha \le \beta$. If further α has all its components equal, P is in $\mathscr{F}^{BS}$ (resp. in $\mathscr{B}^{BS}$) if $\alpha > \beta$ (resp. if $\alpha < \beta$).

This corollary is a straightforward consequence of Theorems 1 and 2, because when $\alpha_m = \beta_m = 1/M$, the search cost ϕ_R does not depend on R [$\phi_R = (M+1)/2$]. So we get examples of processes F with arbitrary stationary law, such that move-to-front is optimal in $\mathscr{C}$ and, in particular, such that the optimal static ordering (and the counter scheme) are more expensive than move-to-front, not only from its heavy memory requirements, but also from its search cost. As a matter of fact, this corollary can be proven directly, observing that for these special matrices P, move-to-front coincides in the long run with the optimal algorithm given in [6], that is, the counter scheme revisited for Markov chains. This is no longer true for the matrices given in Theorem 3.

We have constructed elements of $\mathscr{F}^*$ from other elements of $\mathscr{F}^*$. This has an interest in itself (supporting the notion that optimality of move-to-front "increases" when locality increases), but it would be ineffective if there were almost no elements in $\mathscr{F}^*$. So in the following text we give some significant classes of elements of $\mathscr{F}^*$.

THEOREM 3. *In the following two cases, P belongs to $\mathscr{F}^B$:*

(i) *when S is the union of p classes with q elements each and when*

$$p_{m,n} = \begin{cases} \alpha, & \text{if } m \text{ and } n \text{ are in the same class},\\ \beta, & \text{otherwise} \end{cases}$$

with

$$\alpha \ge \beta;$$

(ii) *when, S being identified in a natural way with $\mathbb{Z}/M\mathbb{Z}$, we have*

$$p_{m,m} = \alpha, \qquad p_{m,m+1} = \beta, \qquad p_{m,m-1} = \gamma$$

with $\alpha + \beta + \gamma = 1$ and when further

$$\alpha \ge (\beta^{M-1} + \gamma^{M-1})(\beta - \gamma)/(\beta^{M-1} - \gamma^{M-1}).$$

When P assumes one of the previously specified forms, the corresponding inequality is also a necessary condition for P to belong to $\mathscr{F}^B$. The corresponding strict inequality implies that P belongs to $\mathscr{F}^{BS}$.

We obtain, in the course of the proof, stationary costs for R_F, agreeing with the formula of Lam, Leung and Siu [19], even if for (ii) the hypotheses required in [19] are not satisfied. We also obtain the value of the overwork v_R, whatever the initial state, up to an additive constant.

Note that the construction of P_α from P appears in [1], pages 103 and 108–109, in the special case where $p_{i,j}$ does not depend on i: It is considered an adequate model for real reference strings. According to the authors, "there is *adequacy in the weak sense* if the model reflects the two properties of locality and rare references," "the *locality property* is exhibited when a program tends to use small subsets of its pages for relatively long periods of time" and "a program has the *property of rare references* if, in spite of locality sets, the reference strings contain rarely used pages." For kernels P of

Theorem 3, the frequency of references to any given item is $1/M$, because P admits a transitive group of symmetries [5]. As a consequence, these kernels do not satisfy the property of rare references. However, Theorem 1 allows us to exhibit elements of $\mathscr{F}^*$ showing adequacy in the weak sense. Let P be an element of $\mathscr{F}^*$ as described in Theorem 3(i). The associated P_α reflects both properties, provided that $\min_i \alpha_i$ is small with respect to $\max_i \alpha_i$, because according to P_α, the frequency of item i is proportional to $1/\alpha_i$.

It may seem quite arbitrary to be so fixed on the optimality of R_F and R_B, rather than any other rule. Theorem 4 should dissipate this feeling, to some extent. In the sequel, a rule of $\mathscr{C}$ will be called *constant* if it belongs to $\mathscr{C}_L$ and if the associated mapping (from $\mathscr{P}$ to A) is constant. There are thus M constant rules, and among them are R_F and R_B.

THEOREM 4. *If the class of Bellman-optimal library-type rules is not void, it contains a constant rule.*

I have no example for which a constant rule is optimal in $\mathscr{C}$ if it is not R_F or R_B, but we may expect that, R_F and R_B being some kind of extreme points of the set of constant rules, if a constant rule is optimal, then either R_F or R_B is optimal. Thus a statement such as "*If the class of optimal library-type rules is not void, it contains R_F or R_B*" would be more satisfactory than Theorem 4, but this remains a conjecture.

5. Generalizations. We can replace the standard cost $c((x,k),a) = x(k)$ by the more general cost $c((x,k),a) = \kappa(x(k))$, where κ is any mapping from $\mathscr{P}$ to $\mathbb{R}$, and Theorems 1 and 4 remain true with natural changes to the definition of R_F and R_B: Set κ^* (resp. κ^{**}) for the minimum value (resp. the maximum) of the mapping κ and set m^* (resp. m^{**}) to be a member of $\mathscr{P}$ such that

$$\kappa(m^*) = \kappa^* \quad [\text{resp. } \kappa(m^{**}) = \kappa^{**}].$$

Finally, let R_F (resp. R_B) be the constant rule associated with place m^* (resp. m^{**}).

With these changes, Theorem 2 remains true, replacing $\mathscr{F}^{BS}$ and $\mathscr{B}^{BS}$ with $\mathscr{F}^B$ and $\mathscr{B}^B$. Furthermore, Theorem 2 remains true as it is, assuming that κ has only one extremum of each kind.

Finally we may consider the case when $\mathscr{P}$ is a tree (see [9] and [20]). Then the part of Theorem 2 concerning R_F remains true, provided that m^* is the root of the tree.

6. Preliminary results.

PROPOSITION 6.1. *The stochastic process $X = (X_n)_{n \geq 0}$, where*

$$X_n = (D_n, F_n),$$

is a controlled Markov process with state space $S' = \mathscr{S}_M \times S$ and action set $A = \mathscr{P}$.

PROOF. Let $T_{k,a}x$ be the disposition of the items after access to item k and replacement of k on place a, if the previous disposition were x. Observe that X satisfies the Markov property, because

$$(6.1)\qquad \begin{aligned} &\mathbb{P}[X_{n+1} = (x',k') | X_n = (x,k), A_n = a, \text{and } (A_{n-1}, H_{n-1}) \in B] \\ &\quad = \begin{cases} p_{k,k'}, & \text{if } x' = T_{k,a}(x), \\ 0, & \text{otherwise,} \end{cases} \end{aligned}$$

which does not depend on B. □

Unless otherwise indicated, in the sequel the search cost incurred in state $i = (x,k)$ will be $\kappa(x(k))$, where κ is any mapping from $\mathscr{P}$ to $\mathbb{R}$. We shall consider a new cost function

$$(6.2)\qquad \bar{c}(i,a) = \mathbb{E}[C_1 | X_0 = i \text{ and } A_0 = a] = \sum_{m=1}^{M} p_{k,m}\kappa(T_{k,a}x(m)).$$

Optimal (resp. 1^- optimal, Bellman-optimal, etc.) rules are exactly the same for c and $\bar{c}$, and ϕ_R and y_R are unaffected by this change, because we have (see [5])

$$\overline{\Psi}_R(i,\beta) = (\Psi_R(i,\beta) - \kappa(x(k)))/\beta.$$

When R is in $\mathscr{C}_D$, say $R := (f)$, then we shall set

$$c_R(i) := \bar{c}(i, f(i)) \quad \text{and} \quad p_{i,j}(R) := p_{i,j}(f(i));$$

$P_R = [p_{i,j}(R)]_{i,j \in S'}$ will be the kernel of $(X_n)_{n \geq 0}$ when rule R is used and $\pi_{i,j}(R)$ will be the long run average sojourn time in state j, starting from i:

$$\pi_{i,j}(R) = \lim_{T \to +\infty} \frac{1}{T+1} \sum_{t=0}^{T} p_{i,j}^{(t)}(R).$$

The following classical result will be needed in the sequel (see [8], page 68). Let $\Pi(R)$ be the matrix $[\pi_{i,j}(R)]$. Consider the system

$$(6.3)\qquad \begin{cases} v - P_R v = c_R - \phi_R, \\ \Pi(R)v = 0. \end{cases}$$

PROPOSITION 6.2. *The unique solution of* (6.3) *is* $\nu_R = (\nu_R(i))_{i \in S'}$.

DEFINITION 6.1. Let Φ associate each state with the next disposition of items, according to move-to-front. For $i = (x,m)$ in S', let S_i be the set of states $j = (y,m) \in S'$ such that the relative order of $S - \{m\}$ in i and j is identical [i.e., such that $\Phi(i) = \Phi(j)$; see Figure 2].

LEMMA 6.1. *Let* $p_j = (p_{j,k}(a))_{k \in S', a \in A}$ *and* $c_j = (\bar{c}_{(j,a)})_{a \in A}$. *Then* p_j *and* c_j *are constant on each* S_i.

This means that the future of the process, when its present state is $i = (x,m)$, depends on the relative position of the files of $S - \{m\}$, but it does

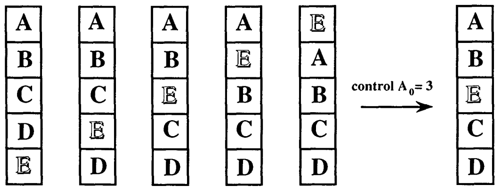

FIG. 2.

not depend on the position of m. For instance, if we assume that $F_0 = E$, the action $A_0 = 3$ gives the same result if D_0 is any of the dispositions in Figure 2 and this is still true for any value of A_0.

PROOF OF LEMMA 6.1. The assertion about p_j follows from

$$T_{m,a} T_{m,b} x = T_{m,a} x$$

and from (6.1). It implies the assertion concerning c_j, considering (6.2). □

Lemma 6.1 points out the class $\mathscr{C}_C$ of rules $R := (f)$ that factor through Φ. Such rules will be called *C-rules*. From Lemma 6.1, it is clear that "the" optimal decision is the same at each state of S_i, so we have the following lemma.

LEMMA 6.2. *Among Bellman-optimal rules, at least one is a C-rule.*

Lemma 6.2 follows at once from Lemma 6.1 and from results about "lumpable" controlled Markov chains stated in Section 3 of [5]. The idea is that $\Phi(X_n)$ is Markov if a C-rule is used, and that consequently ϕ_R, c_R, $\bar{v}_R$ and y_R factor through Φ.

Another useful property of C-rules is the nice behavior of the trajectories of X when we slow the time for F, by adding an arbitrary time of sojourn in each term of the sequence $(F_n)_{n \geq 0}$, before jumping to the next state. Let us drop the random considerations for a while: Let $(F_n)_{n \geq 0}$ be a nonrandom sequence of items and let $(X_n)_{n \geq 0}$ be the corresponding nonrandom sequence of states of the sequential search data structure. Let $(F'_n)_{n \geq 0}$ be a new nonrandom sequence of items built by replacing each F_n by a sequence of Y_n items, each of them equal to F_n [$Y_n \geq 1$; if all Y_n are 1, $(F_n)_{n \geq 0} = (F'_n)_{n \geq 0}$]. Let $(X'_n)_{n \geq 0}$ be the corresponding sequence of states for the sequential search data structure.

LEMMA 6.3. *If R is a C-rule, $(X'_n)_{n \geq 0}$ is built from $(X_n)_{n \geq 0}$ by replacing each X_n by a sequence of Y_n members of S', the first one being X_n and the $Y_n - 1$ remaining terms being all equal to (D_{n+1}, F_n).*

For instance, this is true for rules such as move-to-front or move-to-back, as we see in Figure 3 and 4, where the sequence $(F_n)_{n \geq 0} = DBAB\ldots$ is

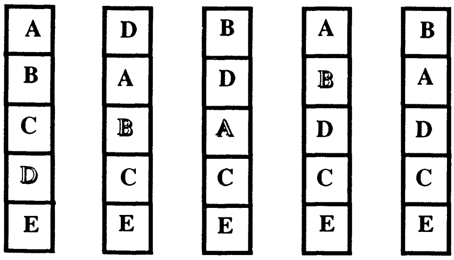

FIG. 3.

replaced with $(F'_n)_{n \geq 0} = DDDBBAAAAAB\ldots$, i.e., $(Y_0, Y_1, Y_2) := (3, 2, 5)$. If move-to-front is used, the sequence $(D_0, D_1, \ldots, D_4)$ is as shown in Figure 3 and through slowing, it becomes $(D'_0, D'_1, \ldots, D'_{11})$ as in Figure 4, where the trajectory and the ending state do not depend on the values of Y_i. One can be convinced, using this kind of figure, that for a rule outside of $\mathscr{C}_C$ (e.g., transpose) the trajectory becomes more strongly dependent on the values of the Y_i's and in a quite unpredictable way.

PROOF OF LEMMA 6.3. From the figures, the lemma is clear. The key point is that, if $F_{n+1} = F_n$, then X_{n+1} is in the same S_i as X_n, so if, furthermore, R is a C-rule, then $A_{n+1} = A_n$. Consequently, X_{n+k} remains equal to X_{n+1} as long as F_{n+k} remains equal to F_n. □

It is a well known fact that if the foregoing Y_n are taken at random with the law

$$\mathbb{P}(Y_n = k/F_n = m) := (1 - \alpha_m)^{k-1}\alpha_m, \qquad k \geq 1,$$

then F'_n is a Markov chain with transition matrix P_α. More precisely, let $(Z_n)_{n \geq 0}$ be a sequence of i.i.d. random variables, independent of $(F_n)_{n \geq 0}$, in which $Z_n = (Z_{n,m})_{1 \leq m \leq M}$ and the $Z_{n,m}$ are mutually independent with laws, respectively,

$$\mathbb{P}(Z_{n,m} = k) := (1 - \alpha_m)^{k-1}\alpha_m, \qquad k \geq 1,$$

and take

$$Y_n := Z_{n, F_n}.$$

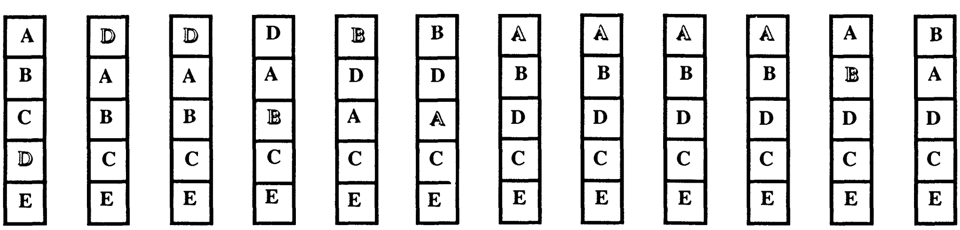

FIG. 4.

Then F_n' is a Markov chain with transition matrix P_α. Consider a C-rule (f) and let $\phi_f(i, \alpha)$ be its stationary search cost when the starting state is i and when the transition matrix of the item process $(F_n')_{n \geq 0}$ is P_α. Finally, let π_n be the stationary probability of the item n.

LEMMA 6.4. *Assume that P has only one ergodic class. Then we have*

$$\phi_f(i, \alpha) = \frac{\phi_f(i) + \sum_{(y,n) \in S'} \pi_{i,(y,n)}(f)(1/\alpha_n - 1)\kappa(f(y,n))}{\sum_{n=1}^{M}(\pi_n/\alpha_n)},$$

provided that no α_m is zero.

PROOF. We give the proof under the additional assumption that $X = (X_n)_{n \geq 0}$ has only one ergodic class. In the general case the idea is similar, but needs heavy notation, because we have to consider each typical trajectory for X and Y before going to expectations. Assume the process X starts at a given recurrent state $i = (x, m)$ and let T be the time of the first return of X to i. Then the time of first return of X' to i is given by

$$T' = \sum_{t=0}^{T-1} Y_{t, F_t},$$

provided that $x(m) \neq f(x, m)$. The cumulated search cost before T' is similarly given by

$$C' = \sum_{t=0}^{T-1} \left[C_t + \kappa(f(X_t))(Y_{t, F_t} - 1)\right].$$

Because $\phi_f(i, \alpha) = \mathbb{E}(C')/\mathbb{E}(T')$, we have to compute the two last expectations, and we obtain

$$\mathbb{E}(T') = \sum_{(y,n) \in S'} \frac{\pi_{i,(y,n)}(f)}{\pi_{i,i}(f)} \frac{1}{\alpha_n}$$

because if X sojourns N_j units of time in state $j = (y, n)$ between two consecutive passages at i, then the corresponding elapsed time for X' will be the sum of N_j i.i.d. random variables (independent from N_j) with mean $1/\alpha_n$, and because the expectation of N_j is well known to be $\pi_{i,j}/\pi_{i,i}$. Similar considerations give

$$\mathbb{E}(C') = \sum_{(y,n) \in S'} \frac{\pi_{i,(y,n)}(f)}{\pi_{i,i}(f)} \left[\kappa(y(n)) + \left(\frac{1}{\alpha_n} - 1\right)\kappa(f(y,n))\right].$$

The existence of a recurrent state (x, m) satisfying $x(m) \neq f(x, m)$ is insured by the additional assumption. □

7. Proof of Theorem 2. In this section, because the only rule of interest is move-to-front, we shall denote this last rule by R. When λ is 0, it is easy to compute directly $\Psi_R(\beta, i, \alpha)$. One finds that the stationary search cost

$\phi_R(i, \alpha)$ is κ^* and that $\bar{v}_R(i, \alpha)$ is zero, in which $\phi_R(i, \alpha)$ and $\bar{v}_R(i, \alpha)$ are related to P_α. So Theorem 2 holds true in this case and we can assume in the sequel that λ is positive. We need to compute $y_R(i, \alpha)$, that is, $p_{i,j}(a, \alpha)$, $c_R(i, \alpha)$, $\phi_R(i, \alpha)$ and $\bar{v}_R(i, \alpha)$ in terms of the corresponding values for $\lambda = 1$, in order to check Bellman's condition. We have from (6.1) that

$$p_{(x,m),(y,n)}(a, \alpha) = \lambda p_{(x,m),(y,n)}(a),$$

except when $y = T_{m,a}x$ and $n = m$, where

$$p_{(x,m),(y,n)}(a, \alpha) = 1 - \lambda + \lambda p_{(x,m),(y,n)}(a).$$

It follows at once that

$$c_R(\alpha, i) = \lambda c_R(i) + (1-\lambda)\kappa^*. \tag{7.1}$$

Lemma 6.4 gives

$$\phi_R(\alpha, i) = \lambda \phi_R(i) + (1-\lambda)\kappa^*. \tag{7.2}$$

We claim that

$$\bar{v}_R(\alpha) = \bar{v}_R \tag{7.3}$$

and, assuming (7.3), we can compute $y_R(\alpha) := (y_R^1(\alpha), y_R^2(\alpha))$ in terms of y_R:

$$\begin{aligned} y_R^1(i, a, \alpha) &= \sum_{j \in S'} p_{i,j}(a, \alpha)\phi_R(j, \alpha) - \phi_R(i, \alpha) \\ &= \lambda\Big(\lambda \sum_{j\in S'} p_{i,j}(a)\phi_R(j) + (1-\lambda)\phi_R((T_{m,a}x, m)) - \phi_R(i)\Big), \end{aligned}$$

in which we let i be (x, m). Because $(T_{m,a}x, m)$ and i belong to the same S_k, we get

$$y_R^1(i, a, \alpha) = \lambda^2 y_R^1(i, a). \tag{7.4}$$

Similarly,

$$\begin{aligned} y_R^2(i, a, \alpha) &:= \bar{c}(i, a, \alpha) + \sum_{j\in S'} p_{i,j}(a, \alpha)\bar{v}_R(j, \alpha) - \bar{v}_R(i, \alpha) - \phi_R(i, \alpha) \\ &= (1-\lambda)(\kappa(a) - \kappa^*) + \lambda \bar{c}(i, a) + \lambda \sum_{j \in S'} p_{i,j}(a)\bar{v}_R(j) \\ &\quad + (1-\lambda)\bar{v}_R((T_{m,a}x, m)) - \bar{v}_R(i) - \lambda\phi_R(i) \end{aligned}$$

and because $\bar{v}_R$ factors through Φ,

$$y_R^2(i, a, \alpha) = (1-\lambda)(\kappa(a) - \kappa^*) + \lambda y_R^2(i, a). \tag{7.5}$$

So we have proved that move-to-front satisfies Bellman's condition with respect to P_α. Furthermore, if κ has value κ^* only at point m^* and if λ is not 1, then P_α belongs to $\mathscr{F}^{BS}$.

When we deal with move-to-back instead of move-to-front (assuming that it is still denoted by R), then (7.3), (7.1), (7.2), (7.5) and (7.4) still hold true, κ^* being eventually replaced with κ^{**}. □

PROOF OF (7.3). Consider the new Markov chain $W = (W_n)_{n \ge 0}$ defined by

$$W_n = \Phi(X_n)$$

and let $\tilde{P}$ and $\tilde{P}_\alpha$ be the kernels of W corresponding to P_R and $P_R(\alpha)$. We have

$$\tilde{P}_\alpha = (1 - \lambda)I + \lambda \tilde{P},$$

and, as a consequence, $\tilde{\Pi} = \tilde{\Pi}_\alpha$. Because c_R factors through Φ, we can set

$$c_R(\alpha, i) = \tilde{c}(\alpha, \Phi(i)), \qquad \bar{v}_R(\alpha, i) = \tilde{v}(\alpha, \Phi(i)) \quad \text{and}$$
$$\phi_R(\alpha, i) = \tilde{\phi}(\alpha, \Phi(i)).$$

Now, according to Proposition 6.2 and relations (7.1) and (7.2),

$$\left(I - \tilde{P}_\alpha\right)\tilde{v} = \lambda(I - \tilde{P})\tilde{v} = \lambda\left(\tilde{c} - \tilde{\phi}\right) = \tilde{c}(\alpha) - \tilde{\phi}(\alpha)$$

and

$$\tilde{\Pi}_\alpha \tilde{v} = \tilde{\Pi}\tilde{v} = 0.$$

Proposition 6.2 thus implies that $\tilde{v} = \tilde{v}(\alpha)$. □

8. Proof of Theorem 1. In this proof, in contrast to that of Theorem 2, it is no longer sufficient to study the consequences of replacing P with P_α for the performance of move-to-front, because move-to-front does not necessarily meet Bellman's condition under P. The only way we can prove optimality of move-to-front under P_α seems to be by comparing the performance of move-to-front under P_α to the performance of *each* rule of $\mathscr{C}_D$ under P_α. Fortunately, Lemma 6.2 allows us not to have to study the consequences of the change of P to P_α for each rule of $\mathscr{C}_D$ (a task that would be very hard with regard to the rule "transpose," for instance), but only for C-rules. Fortunately again, it turns out that, for C-rules, this task is "doable" (see Lemma 6.4).

STEP 1. *F has only one ergodic class and the α_m are positive.* We assume that P is in $\mathscr{F}^*$. We have to show that

$$\phi_{R_F}(i, \alpha) \le \phi_f(i, \alpha) \tag{8.1}$$

for any stationary rule (f), under the hypothesis that $\phi_{R_F}(i) \le \phi_f(i)$. Keeping in mind Lemma 6.2, we can restrict ourselves to C-rules and (8.1) becomes

$$\frac{\phi_{R_F}(i) + \sum_{(y,n)\in S'} \pi_{i,(y,n)}(R_F)(1/\alpha_n - 1)\kappa^*}{\sum_{n=1}^{M} \pi_n(1/\alpha_n)}$$
$$\le \frac{\phi_f(i) + \sum_{(y,n)\in S'} \pi_{i,(y,n)}(f)(1/\alpha_n - 1)\kappa(f(y,n))}{\sum_{n=1}^{M} \pi_n(1/\alpha_n)}.$$

Then, (8.1) follows from

$$\sum_{(y,n)\in S} \pi_{i,(y,n)}(R_F)\left(\frac{1}{\alpha_n}-1\right)\kappa^* = \sum_{(y,n)\in S} \pi_{i,(y,n)}(f)\left(\frac{1}{\alpha_n}-1\right)\kappa^*.$$

Let us proceed to the proof of the second part of the theorem, concerning $\mathscr{B}^*$. Because α is now an element of $[1, +\infty[^M$, a probabilistic construction of $Q = P_\alpha$ in terms of a Markov chain with transition matrix P is no longer available. Therefore, we have to reverse our point of view: Set $\beta = (1/\alpha_1, 1/\alpha_2, \dots, 1/\alpha_M)$, so that β is in $[0,1]^M$ and so that we have a probabilistic construction of P as Q_β. The hypothesis is now

$$\phi_{R_F}(i,\beta) \le \phi_f(i,\beta) \tag{8.2}$$

for any $R := (f)$ in $\mathscr{C}_D$. The argument β in these expressions means that the value is related to Q_β (i.e., P) and we drop β when β is $(1,1,\dots,1)$. We want to deduce from (8.2) that

$$\phi_{R_F}(i) \le \phi_f(i) \tag{8.3}$$

for any C-rule f. Now (8.2) can be written

$$\phi_{R_F}(i) + \sum_{(y,n)\in S} \pi_{i,(y,n)}(R_B)\left(\frac{1}{\beta_n}-1\right)\kappa^{**}$$
$$\le \phi_f(i) + \sum_{(y,n)\in S} \pi_{i,(y,n)}(f)\left(\frac{1}{\beta_n}-1\right)\kappa(f(y,n)),$$

and it is quite clear that

$$\sum_{(y,n)\in S} \pi_{i,(y,n)}(R_B)\left(\frac{1}{\beta_n}-1\right)\kappa^{**} \ge \sum_{(y,n)\in S} \pi_{i,(y,n)}(f)\left(\frac{1}{\beta_n}-1\right)\kappa(f(y,n)),$$

so (8.3) follows at once.

STEP 2. *F has several ergodic classes and the α_m are positive.* Let $E_1, E_2, \dots, E_k$ be the ergodic classes of F and let E be the union of these classes. Furthermore, let E' (resp. E'_r) be the set of states $i = (x,m)$ such that m is in E (resp. E_r). If i is in E', Lemma 6.4 still holds true and (8.1) follows as in Step 1. Let T (resp. T') be the time of first entrance of F in E (resp. F' in E) and set

$$p^*_{i,j}(f) := \mathbb{P}_{i,f}(X'_{T'} = j)$$

for a given C-rule (f). When i is not in E', we have

$$\phi_f(\alpha,i) = \sum_{r=1}^{k} \sum_{j\in E'_r} p^*_{i,j}(f)\phi_f(\alpha,j). \tag{8.4}$$

Clearly, we have that (i) $\sum_{j\in E'_r} p^*_{i,j}(f)$ does not depend on f, because if $i = (x,m)$, we have

$$\sum_{j\in E'_r} p^*_{i,j}(f) = \mathbb{P}_m(F'_{T'} \in E_r).$$

We claim that (ii) $j \to \phi_{R_F}(\alpha, j)$ is constant on each E'_r. Assuming this last point, we can give the end of the proof of Step 2. With $\phi_{R_F}(\alpha, r)$ standing for the value of $\phi_{R_F}(\alpha, j)$ when j belongs to E'_r,

$$\begin{aligned}\phi_{R_F}(\alpha, i) &= \sum_{r=1}^{k} \phi_{R_F}(\alpha, r) \sum_{j \in E'_r} p^*_{i,j}(R_F) \\ &= \sum_{r=1}^{k} \sum_{j \in E'_r} p^*_{i,j}(f) \phi_{R_F}(\alpha, j).\end{aligned} \tag{8.5}$$

Comparing (8.4) and (8.5) ends the proof, because (8.1) holds true when i belongs to E'.

Now, let us prove point (ii). Unfortunately, there are several ergodic classes in E'_r (which is closed). Let Φ_1 be defined on E'_r by $\Phi_1(x, m) := (x|_{E_r}, m)$. One easily sees that $U = (U_n)_{n \geq 0} := (\Phi_1(X'_n))_{n \geq 0}$ is a Markov chain, and that because items out of E'_r are never required, the cost C_n incurred at time n depends only on U_n. Using an argument similar to that in Letac [20], page 18, or in Lam, Leung and Siu [19], Theorem 2.1.2, we see that U has a regeneration state and therefore a unique ergodic class.

The part of Theorem 1 concerning $\mathscr{B}^*$ can be proven in the same way.

STEP 3. *Some α_m are zero.* Let $\mathscr{E}$ be the set of items m such that α_m is zero and let $\mathscr{E}'$ be the set of states $i = (x, m)$ such that m belongs to $\mathscr{E}$. There are two subclasses among ergodic classes of F': The ergodic classes of F that do not intersect $\mathscr{E}$ and the classes $\{m\}$ in which m belongs to $\mathscr{E}$. Let $E_1, E_2, \dots, E_k$ be the ergodic classes of the first kind, the union of ergodic classes of the second kind (say $E_{k+1}, E_2, \dots, E_{k'}$) being $\mathscr{E}$, and let E'_r be defined from E_r as in Step 2. Relation (8.1) still holds true for i in the E'_r of the first kind, with the same proof as in Step 1. For a state i belonging to $\mathscr{E}'$, (8.1) holds true because

$$\phi_{R_F}(i, \alpha) = \kappa^*.$$

The proof is finished, with the help of points (i) and (ii), as in the second step. □

9. Proof of Theorem 3. In this section we consider only the cost function $\kappa_0(p) = p$ that is commonly considered in the related literature. From [5], we know two things: On the one hand, it is more likely that a library-type rule is optimal if the kernel P of F has a transitive group of symmetries and, on the other hand, (6.3) is easier to solve, due to some state space reduction for X, if P is state-transitive. Searching for kernels P of the class $\mathscr{F}^B$, we shall thus focus on state-transitive kernels. The two examples that follow are kernels of random walks for some group structure on S.

A. Locality by bunches. Now we shall prove part (i) of Theorem 3. Let $M = pq$ and let $(K_1, K_2, \dots, K_q)$ be a partition of S with $\#K_i = p$ for each i;

let $p\alpha + (q-1)p\beta = 1$. When a file of K_i is required, let the next file be any of the files in K_i with probability α and any of the files outside K_i with probability β. When $\alpha > \beta$, this is a reasonable model for the locality phenomenon, as pictured in [17], page 399: "small groups of keys tend to occur in bunches." With respect to the locality assumption, the condition that the K_i have the same cardinality is artificial, but this condition is required for the state-transitivity of P.

PROOF OF THEOREM 3(i). When $\beta = 0$, the optimality of the move-to-front rule is easy to check directly and the minimal cost is $(p+1)/2$, so we assume, from now on, that β is positive. This insures ergodicity of X under move-to-front. From [25], it follows that the set of solutions of

$$v_i - (P_R v)_i = c_R(i) - \phi, \qquad i \in S', \tag{9.1}$$

with unknowns $(v_i)_{i \in S'}$ and the real number ϕ, is exactly $\{((v_R(i) + K)_{i \in S'}, \phi_R) | K \in \mathbb{R}\}$. Clearly y_R remains unchanged if v_R is replaced with $v_R + K$, so ergodicity allows us to replace the solution of (6.3), which requires preliminary computation of ϕ_R and Π_R, with the solution of (9.1). Another consequence of ergodicity on Bellman's condition is that y_R^1 is identically zero.

In this section i denotes the state (x, m) and $K(m)$ denotes, among subsets K_q, the one that contains item m. Symmetries of P (see [5]) and the probabilistic interpretation of $\bar{v}_R$ as a starting cost suggest the following trial for the solution of (9.1):

$$v_i = \lambda \frac{1}{p} \sum_{n \in K(m)} x'(n), \tag{9.2}$$

in which λ is positive and x' denotes $T_{m,1}x$. Let us take

$$G(x, A) = \frac{1}{\#A} \sum_{n \in A} x(n),$$

the center of gravity of the subset A of S, when the disposition is x. Assuming (9.2), let us check that there exists a real number ϕ such that $((v_i)_{i \in S'}, \phi)$ is a solution of (9.1), or equivalently, of

$$v_{(x,m)} = \bar{c}((x,m),1) - \phi + \sum_{n \in K(m)} \alpha v_{(x',n)} + \sum_{n \notin K(m)} \beta v_{(x',n)}. \tag{9.3}$$

Our guess will be right if there exist ϕ and λ such that, independently of x and m, we have

$$\begin{aligned} \lambda G(x', K(m),) = \bar{c}((x,m),1) - \phi + \lambda\alpha \sum_{n \in K(m)} G(T_{n,1}x', K(n),) \\ + \lambda\beta \sum_{n \notin K(m),} ,G(T_{n,1}x', K(n),). \end{aligned} \tag{9.4}$$

The result follows from the next lemma

LEMMA 9.1. *Let x be a disposition, A a subset of F and U a random element of A. Then*

$$\mathbb{E}[G(T_{U,1}x, A)] = \frac{\#A - 1}{\#A} G(x, A) + \frac{1}{\#A} \frac{\#A + 1}{2}.$$

PROOF.

$$G(T_{U,1}x, A) - G(x, A) = \frac{1}{\#A}[1 - x(U) + r(U) - 1] \leq 0,$$

in which $r(U)$ is the rank of U in A,

$$r(U) = \#\{m \in A | x(m) \leq x(U)\},$$

$r(U) - 1$ is the number of files in A whose positions are increased by 1 when U goes to position 1 and $1 - x(U)$ is the contribution to $G(T_{U,1}x, A) - G(x, A)$ of the move of U. Finally, it is easy to check that

$$\mathbb{E}[x(U)] = G(x, A) \quad \text{and} \quad \mathbb{E}[r(U)] = \frac{\#A + 1}{2}. \qquad \square$$

LEMMA 9.2. *For any disposition x,*

$$\frac{1}{q} \sum_{i=1}^{q} G(x, K_i) = \frac{M + 1}{2}.$$

LEMMA 9.3. *For any disposition x,*

$$\bar{c}((x, n), a) = (\alpha - \beta) p G(T_{n,a}x, K(n)) + \beta p q \frac{M + 1}{2}.$$

PROOF OF LEMMA 9.3 Lemma 9.2 is clear and gives

$$\begin{aligned}\bar{c}((x, n), a) &= \alpha p G(T_{n,a}x, K(n)) + \beta p \sum_{K_i \neq K(n),} G(T_{n,a}x, K_i) \\ &= (\alpha - \beta) p G(T_{n,a}x, K_n) + \beta p q \frac{M + 1}{2}. \qquad \square\end{aligned}$$

Now, by applying Lemma 9.1 with $A = K(m)$, we can simplify each term of (9.4):

$$\lambda \alpha \sum_{n \in K(m)} G(T_{n,1}x', K(n)) = \lambda \alpha p \left[\frac{p - 1}{p} G(x', K(m)) + \frac{1}{p} \frac{p + 1}{2}\right]$$

and

$$\lambda \beta \sum_{n \notin K(m),} G(T_{n,1}x', K(n),) = \lambda \beta p \sum_{K_i \neq K(m),} \left[\frac{p - 1}{p} G(x', K_i) + \frac{1}{p} \frac{p + 1}{2}\right],$$

so that

$$\lambda\alpha \sum_{n\in K(m)} G(T_{n,1}x', K(n)) + \lambda\beta \sum_{n\notin K(m),} G(T_{n,1}x', K(n))$$
$$= \lambda(\alpha-\beta)p\left[\frac{p-1}{p}G(x', K(m)) + \frac{1}{p}\frac{p+1}{2}\right]$$
$$+ \lambda\beta p \sum_{i=1}^{q}\left[\frac{p-1}{p}G(x', K_i) + \frac{1}{p}\frac{p+1}{2}\right]$$
$$= \lambda(\alpha-\beta)(p-1)G(x', K(m)) + \lambda\psi(\alpha, M, p),$$

in which

$$\psi(\alpha, M, p) = (\alpha-\beta)\frac{p+1}{2} + \beta M\left(\frac{p-1}{p}\frac{M+1}{2} + \frac{1}{p}\frac{p+1}{2}\right).$$

Using Lemma 9.3 and putting the pieces together, we reduce the system (9.3) to

$$\begin{aligned} \lambda &= (\alpha-\beta)p + \lambda(\alpha-\beta)(p-1), \\ \phi &= \beta p q\frac{M+1}{2} + \lambda\psi(\alpha, M, p) \end{aligned} \tag{9.5}$$

and because (9.5) has a solution, so does (9.3). When $\beta > 0$, X is irreducible, so we can use any solution of (9.1) in Bellman's condition. Furthermore, we have

$$\lambda = (\alpha-\beta)\frac{p}{\alpha + (M-1)\beta}.$$

We now have to check that

$$1 \in \operatorname*{argmin}_{a}\left\{\bar{c}((x,m),a) + \sum_{(y,n)\in S'} p_{(x,m),(y,n)}(a)v_{(y,n)}\right\}.$$

Computations similar to those given above show that

$$\bar{c}((x,m),a) + \sum_{(y,n)\in S'} p_{(x,m),(y,n)}(a)v_{(y,n)}$$
$$= \lambda G(T_{m,a}x, K(m)) + \beta p q\frac{M+1}{2} + \lambda\psi(\alpha, M, p).$$

Because λ has the same sign as $\alpha - \beta$, Theorem 3(i) follows from the fact that $G(T_{n,a}x, A)$ is an increasing function of a when n belongs to A. □

Note that (9.5) easily gives the explicit expression of the minimal stationary search cost ϕ, in that context.

B. Nearest neighbor random walks on cyclic groups. For Theorem 3(ii) we identify S with $\mathbb{Z}_M$ and we let P be the kernel of the random walk with law

$\mu = \alpha\delta_0 + \beta\delta_1 + \gamma\delta_{-1}$. When $\alpha = 1$, move-to-front is optimal because its stationary search cost is 1, so we assume, from now on, that $\alpha < 1$.

PROOF OF THEOREM 3(ii). It is easy to check that state

$$i = \left(\begin{pmatrix} 1 & 2 & 3 & \cdots & M \\ 1 & M & M-1 & \cdots & 2 \end{pmatrix}, 1 \right)$$

is a regeneration state for X when $\beta > 0$ and $\gamma = 0$, and thus, a fortiori, when $\gamma > 0$. We are thus allowed to use the solution of system (9.1) in y_R. Because $\tau_k(m) = m + k$ is a symmetry of P, we have, from Theorem 2 of [5], that

$$v_{(x,k)} = v_{(\tau_{k-1}x,1)} \quad \text{with } (\tau_k x)(m) = x(m+k). \tag{9.6}$$

From the interpretation of v_R as a starting cost, we have found, when $\gamma = 0$, that $v_R(x,1)$ is the number of inversions of x (see [6]). When $\gamma \neq 0$, a natural idea is thus to try

$$v_{(x,1)} = \sum_{1 \le i < j \le M} \alpha_{i,j} \mathbb{1}_{x(i) > x(j)}$$

as a solution for (9.1). From (9.6), the solution of (9.1) can be reduced to checking that there exists a real number ϕ and coefficients $\alpha_{i,j}$ such that for any x,

$$v_{(x,1)} = x(1) - \phi + \alpha v_{(x',1)} + \beta v_{(\tau_1 x',1)} + \gamma v_{(\tau_{-1}x',1)}, \tag{9.7}$$

where x' still denotes $T_{1,1}x$. Let

$$z_{i,j}(x) = \mathbb{1}_{x(i) > x(j)}.$$

We have

$$z_{i,j}(\tau_1 x) = z_{i+1,j+1}(x), \quad \text{if } 2 \le j < M,$$

$$z_{i,M}(\tau_1 x) = 1 - z_{1,i+1}(x)$$

and a similar statement holds for $z_{i,j}(\tau_{-1}x)$. Furthermore,

$$z_{i,j}(x') = z_{i,j}(x), \quad \text{if } 2 \le i \le M-1,$$

$$z_{1,j}(x') = 0.$$

It follows that (9.7) can be written in terms of the $z_{i,j}(x)$, and that a sufficient condition for (9.7) to hold true is that the coefficients of the $z_{i,j}$, as well as the constant term, are zero. After straightforward computations, we obtain that $(\alpha_{i,i+k})_{1 \le i \le M-k}$ can be written $(w_i)_{1 \le i \le M-k}$, in which $(w_i)_{1 \le i \le M-k+1}$ would be a solution of

$$\gamma w_{i+1} - (\beta + \gamma) w_i + \beta w_{i-1} = 0,$$

with boundary conditions $w_1 = 1$ and $w_{M-k+1} = -1$. So, when $\beta \neq \gamma$,

$$\alpha_{i,j} = 1 - 2\frac{1 - (\beta/\gamma)^{i-1}}{1 - (\beta/\gamma)^{M-j+i}}, \qquad 1 \le i < j \le M,$$

(9.8)

$$\phi = 1 + (M-1)\gamma + \beta \sum_{i=1}^{M-1} \alpha_{i,M}$$

and when $\beta = \gamma$,

$$\alpha_{i,j} = \frac{i+j-M-2}{j-i-M}, \qquad 1 \le i < j \le M, \tag{9.9}$$
$$\phi = 1 + (1-\alpha)\sum_{i=1}^{M-1}\frac{1}{i}.$$

The Bellman condition gives optimality of move-to-front if for any x,

$$1 \in \operatorname*{argmin}_a \left\{ \sum_{(y,n)\in S'} p_{(x,1),(y,n)}(a) v_{(y,n)} \right\}.$$

Let $T_{1,a}x = x_a$. We have

$$\sum_{(y,n)\in S'} p_{(x,1),(y,n)}(a) v_{(y,n)} = \alpha v_{(x_a,1)} + \beta v_{(x_a,2)} + \gamma v_{(x_a,M)}.$$

By ignoring terms that depend on x but not on a, we only have to check that, for any x,

$$1 \in \operatorname*{argmin}_a \psi_x(a), \tag{9.10}$$

with

$$\psi_x(a) = \sum_{j=2}^{M} \left(\alpha\alpha_{1,j} - \beta\alpha_{j-1,M} + \gamma\alpha_{2,j+1}\right) z_{1,j}(x_a),$$

in which $\alpha_{i,M+1}$ is assumed to be defined by (9.8) or (9.9), and thus to be -1. Let us set

$$\xi_x(a) = \psi_x(a) - \psi_x(a-1), \qquad 2 \le a \le M,$$

and

$$\delta(j) = \alpha\alpha_{1,j} - \beta\alpha_{j-1,M} + \gamma\alpha_{2,j+1}, \qquad 2 \le j \le M.$$

If $x(1)$ is k, we have

$$\begin{aligned} (\xi_x(a))_{2\le a\le M} = \big(&\delta(x^{-1}(1)), \delta(x^{-1}(2)), \dots, \delta(x^{-1}(k-1)), \\ &\delta(x^{-1}(k+1)), \dots, \delta(x^{-1}(M))\big). \end{aligned} \tag{9.11}$$

From (9.11), in order that (9.10) hold true for a given x, it will thus be sufficient that

$$\delta(j) \text{ is nonnegative for each } j. \tag{9.12}$$

However, in order that (9.10) hold true for any x, it is necessary that $\xi_x(2)$ [i.e., $\delta(x^{-1}(1))$] be nonnegative for each x, which is equivalent to (9.12). Condition (9.12) is thus equivalent to the Bellman condition. Straightforward computations prove that

$$\delta(2) = \delta(M) = \min\{\delta(j); 2 \le j \le M\}$$

and that

$$\delta(2) \ge 0 \Leftrightarrow \alpha \ge (\beta^{M-1} + \gamma^{M-1})(\beta - \gamma)/(\beta^{M-1} - \gamma^{M-1}). \qquad \square$$

10. Proof of Theorem 4. Let R be a library-type Bellman-optimal rule and let f be the map such that $R(x, m) = f(x(m))$. From Lemma 6.1 we know that the vectors $(\sum_{k \in S'} p_{j,k}(a)\phi_R(k))_{a \in A}$ and $(\bar{c}(j, a) + \sum_{k \in S'} p_{j,k}(a)\bar{v}_R(k))_{a \in A}$ are constant on the class S_i. We have

$$\phi_R(j) = \min\left\{ \sum_{k \in S'} p_{j,k}(a)\phi_R(k) | a \in A \right\}.$$

The second term of the preceding relation is constant for j in S_i. Thus, and though R is not a C-rule, $\phi_R(j)$ factors through Φ. Similarly $\bar{v}_R(j)$ is constant on S_i and so is the set B_j appearing in the Bellman condition. Because R is optimal, for $j = (y, k)$, B_j contains $f_{(y(k))}$. Let i be (x, m). Because B_j is constant on S_i, B_i contains $f_{(y(m))}$ for any (y, m) in S_i, that is, it contains

$$\{f(T_{m,a}x(m)) | a \in A\} = \{f(a) | a \in \mathscr{P}\} = f(\mathscr{P}).$$

Let a_0 be any fixed element in $f(\mathscr{P})$ and let R' be the library-type policy associated with f', where f' is constant with value a_0. Then Proposition 3.1 implies that the stationary search cost of R' is lower than that of R and is thus minimal. □

Acknowledgments. The author thanks Laurent Alonso, Philippe Bougerol, Philip Feinsilver, Edward M. Reingold, René Schott and an anonymous referee for many helpful comments concerning this paper.

REFERENCES

[1] AVEN, O. I., COFFMAN, E. G. and KOGAN, Y. A. (1987). *Stochastic Analysis of Computer Storage*. Reidel, Dordrecht.

[2] BENTLEY, J. L. and MCGEOCH, C. (1985). Amortized analyses of self-organizing sequential search heuristics. *Comm. ACM* **28** 404–411.

[3] BITNER, J. R. (1979). Heuristics that dynamically organize data structures. *SIAM J. Comput.* **8** 82–110.

[4] CHASSAING, P. (1991). Processus de Markov transitifs et marches aléatoires sur les groupes. *C. R. Acad. Sci. Paris Ser. I* **312** 291–296.

[5] CHASSAING, P. (1992). State-transitive Markov processes, random walks on groups, and self-organizing data structures. Preprint.

[6] CHASSAING, P. (1992). On the optimality of the move-to-front heuristic in case of locality of references. Preprint.

[7] CHUNG, F. R. K., HAJELA, D. J. and SEYMOUR. P. D. (1988). Self-organizing sequential search and Hilbert's inequalities. *J. Comput. Syst. Sci.* **36** 148–157.

[8] DERMAN, C. (1970). *Finite State Markovian Decision Processes. Mathematics in Science and Engineering* **67**. Academic, New York.

[9] DIES, J. E. (1981). *Chaînes de Markov sur les Permutations. Lecture Notes in Math.* **1010**. Springer, Berlin.

[10] FLAJOLET, P., GARDY, D. and THIMONIER, L. (1989). *Random Allocation and Probabilistic Languages. Lecture Notes in Computer Science* **317** 239–253. Springer, New York.

[11] GONNET, G. H., MUNRO, J. I. and SUWANDA, H. (1981). Exegesis of self-organizing linear search. *SIAM J. Comput.* **10** 613–637.

[12] HENDRICKS, W. J. (1976). An account of self-organizing systems. *SIAM J. Comput.* **5** 715–723.

[13] HOFRI, M. and SHACHNAI, H. (1991). Self-organizing lists and independent references: A statistical synergy. *J. Algorithms* **12** 533–555.

[14] HOFRI, M. and SHACHNAI, H. (1991). On the optimality of the counter scheme for dynamic linear lists. *Inform. Process. Lett.* **37** 175–179.

[15] KAN, Y. C. and ROSS, S. M. (1980). Optimal list order under partial memory constraints. *J. Appl. Probab.* **17** 1004–1015.

[16] KAPOOR, S. and REINGOLD, E. M. (1991). Stochastic rearrangement rules for self-organizing data structures. *Algorithmica* **6** 278–291.

[17] KNUTH, D. E. (1973). *The Art of Computer Programming* (*T.3*: *Sorting and Searching*). Addison-Wesley, Reading, MA.

[18] KONNEKER, L. K. and VAROL, Y. L. (1981). A note on heuristics for dynamic organization of data structures. *Inform. Process. Lett.* **12** 213–216.

[19] LAM, K., LEUNG, M. and SIU, M. (1984). Self-organizing files with dependent accesses. *J. Appl. Probab.* **21** 343–359.

[20] LETAC, G. (1978). *Chaînes de Markov sur les Permutations*. S.M.S. Presses de l'Université de Montréal.

[21] PHELPS, R. and THOMAS, R. (1980). Optimality for a special case of self-organizing schemes. *J. Inform. Optim. Sci.* **1** 80–93.

[22] RIVEST, R. (1976). On self-organizing sequential search heuristics. *Comm. ACM* **19** 63–67.

[23] SLEATOR, D. D. and TARJAN, R. E. (1985). Amortized efficiency of list updates and paging rules. *Comm. ACM* **28** 202–208.

[24] TENENBAUM, A. and NEMES, R. M. (1982). Two spectra of self-organizing sequential search algorithms. *SIAM J. Comput.* **11** 557–566.

[25] TIJMS, H. C. (1986). *Stochastic Modelling and Analysis*: *A Computational Approach*. Wiley, New York.

DÉPARTEMENT DE MATHÉMATIQUES
UNIVERSITÉ DE NANCY 1
B.P. 239
54506 VANDOEUVRE CEDEX
NANCY
FRANCE