P. POKAROWSKI (Warszawa)

DIRECTED FORESTS WITH APPLICATION TO ALGORITHMS RELATED TO MARKOV CHAINS

Abstract. This paper is devoted to computational problems related to Markov chains (MC) on a finite state space. We present formulas and bounds for characteristics of MCs using directed forest expansions given by the Matrix Tree Theorem. These results are applied to analysis of direct methods for solving systems of linear equations, aggregation algorithms for nearly completely decomposable MCs and the Markov chain Monte Carlo procedures.

0. Introduction. This work is devoted to computational problems related to Markov chains (MC) on a finite state space. It is a shorten version of the author's Ph.D. thesis [Po 1]. In Section 1, using some combinatorial structures—directed forests—we present formulas and bounds for such characteristics of MCs as the stationary distribution, mean hitting times and eigenvalues of the transition matrix. These formulas and bounds have the form of rational functions of elements of the transition matrix and follow from the Matrix Tree Theorem. We apply these results to analyse four groups of algorithms.

In Section 2 we study direct methods for computing characteristics of MC which are solutions of systems of linear equations. We generalize Grassmann, Taksar and Heyman's version of the Gaussian elimination method. We give bounds for the entrywise relative error of this algorithm.

Section 3 deals with aggregation algorithms for approximation of the characteristics of perturbed MCs. Such algorithms are used to solve large

and sparse linear systems induced by nearly completely decomposable MCs. In this case we generalize the known algorithms for approximating a stationary distribution to other characteristics of MCs and to nonlinear perturbations.

The last two sections are devoted to the study of Markov chain Monte Carlo algorithms (MCMC). In Section 4 we bound errors of a general class of MCMC methods for estimating integrals. In Section 5 we characterize asymptotic correctness of MCMC algorithms for finding a global minimum.

Due to space constraints, proofs have been omitted. They can be found in [Po 1] and will be published elsewhere.

1. Directed forests and Markov chains

1.1. *Preliminaries.* Let S be a given nonempty finite set and $E \subseteq S \times S$. For simplicity we assume that $S = \{1, \ldots, s\}$. The (directed) *graph* with the *state* set S and the *edge* set E is, by definition, the pair $g := (S, E)$. A pair $g_1 := (S_1, E_1)$ is called a *subgraph* of g if $S_1 \subseteq S$ and $E_1 \subseteq E \cap (S_1 \times S_1)$. A subgraph g_1 is called *spanning* if $S_1 = S$. A *path* from a state i to a state j is, by definition, any finite sequence $i_0 = i, i_1, \ldots, i_k = j$ such that $i_m \in S$ and $(i_m, i_{m+1}) \in E$ for $m = 1, \ldots, k-1$. By a *cycle* we mean a path from i to i.

A spanning subgraph without cycles in which from every state there is at most one outgoing edge is called a spanning *forest*, $f = (S, E_f)$. The set $R \subseteq S$ of states of the forest f from which there is no outgoing edge is called the *root* of f. It is easily seen that the root of f is nonempty and for every state $i \in S \setminus R$ there is only one path from i to R. We denote by $F(R)$ the set of all forests in g with root R. For $i \notin R$ and $j \in R$, we denote by $F_{ij}(R)$ the subset of $F(R)$ consisting of all forests with a path from i to j.

Let $\mathbf{A} = (a_{ij})_{i,j \in S}$ be an $n \times n$ complex matrix. The *weighted graph* induced by $\mathbf{A}$ is, by definition, the matrix $\mathbf{A}$ together with the graph $g(\mathbf{A}) := (S, E)$, where $E = \{(i, j) \in S \times S : a_{ij} \neq 0\}$. The (multiplicative) *weight* of a forest $f = (S, E_f)$ in $g(\mathbf{A})$ is defined to be

$$w(f) := \prod_{(i,j) \in E_f} (-a_{ij})$$

(we set $w((S, \emptyset)) := 1$). The weight of a set F of forests in $g(\mathbf{A})$ is defined

to be

$$w(F) := \sum_{f \in F} w(f) \quad (w(\emptyset) := 0).$$

If $F = F(R)$ for some $R \subseteq S$, we write $w(R)$ instead of $w(F(R))$, because the set R determines the set of all forests with root R. Set $w(i) := w(\{i\})$, $w_{jk}(R, k) := w[F_{jk}(R \cup \{k\})]$ and $w_{kl}(R) := w[F_{kl}(R)]$ for $i \in S$, $j, k \in S \setminus R$, $l \in R$.

A matrix $\mathbf{L} := (l_{ij})_{i,j=1}^s$, $l_{ij} \in \mathbb{C}$, is said to be a *laplacian matrix* if $l_{ii} = -\sum_{j: j \neq i} l_{ij}$ for $i = 1, \ldots, s$. To explain the name we note that such matrices appear in solving partial differential equations with the Laplace operator (see for example [Mo]). Symmetric laplacian matrices are known in combinatorics (e.g. [CvDoSa], [Mo]). Their eigenvalues are used to bound some combinatorial parameters. Laplacian matrices have also been studied in the theory of electrical networks under the name of "admittance matrices" or "Kirchhoff matrices" (see [Che], [Mo]).

Let $(\Omega, \mathcal{F}, \mathbf{Pr})$ be a probability space and $X = (X_t)_{t \geq 0}$ a Markov chain (MC) defined on $(\Omega, \mathcal{F})$ and with state space S. Markov chains are usually introduced by a transition probability matrix $\mathbf{P} = (p_{ij})_{i,j \in S}$ (when time is discrete) or by a generator $\mathbf{Q} = (q_{ij})_{i,j \in S}$ (when time is continuous). Let $\mathbf{I}$ denote the $s \times s$ identity matrix. It is easily seen that the matrices $\mathbf{L}(\mathbf{P}) = \mathbf{I} - \mathbf{P}$ and $\mathbf{L}(\mathbf{Q}) = -\mathbf{Q}$ are laplacian matrices induced by $\mathbf{P}$ and $\mathbf{Q}$.

Many facts we consider here are the same for the discrete and continuous case (see [Io], [KeSn] for more details about MCs). For that reason and for simplicity of notation we will introduce MCs by *Markov chain laplacian matrices* (MC laplacian matrices), i.e. laplacian matrices whose off-diagonal elements are nonnegative.

1.2. *Directed forest expansions for cofactors of a laplacian matrix.* For $U, W \subseteq S$ and an $s \times s$ matrix $\mathbf{A}$, denote by $\mathbf{A}(U \,|\, W)$ the submatrix of $\mathbf{A}$ obtained by deletion of the rows and columns indexed by U and W respectively. The *cofactor* of $\mathbf{A}(U \,|\, W)$ is, by definition, the number

$$C_{\mathbf{A}}(U \,|\, W) := (-1)^{\sum_{i \in U} i + \sum_{j \in W} j} \det \mathbf{A}(U \,|\, W).$$

For simplicity of notation we write $\mathbf{A}_{ij}$ instead of $\mathbf{A}(\{i\} \,|\, \{j\})$ and $\mathbf{A}(U)$ instead of $\mathbf{A}(U \,|\, U)$. Let $\mathbf{e}_s$ and $\mathbf{0}_s$ denote the column vectors in which each component is 1 and 0 respectively.

The following lemma allows one to represent many characteristics of MCs in the form of directed forest expansions, i.e. rational functions of weights of sets of forests in $g(\mathbf{L})$. Without loss of generality we can assume that the states are numbered so that $R = \{s - |R| + 1, \ldots, s\}$, where $|R|$ denotes the cardinality of R.

LEMMA 1.1. *Let* $\mathbf{L}$ *be an* $s \times s$ *laplacian matrix*, $R \subseteq S$ *and* $i, j \notin R$. *Then*:

(1) ([FieSe]) $\det \mathbf{L}(R) = w(R)$;
(2) $C_{\mathbf{L}}(R \cup \{j\} \mid R \cup \{i\}) = w_{ij}(R, j)$.

In the proof of the above lemma we use a general version of the "Matrix Tree Theorem" [Che, prob. 4.16, Cha]. A simple consequence of Lemma 1.1 is the following.

LEMMA 1.2. *Let* $\mathbf{L}$ *be an* $s \times s$ *laplacian matrix and* $R \subseteq S$. *Suppose that* $w(R) \neq 0$. *Then*

$$\mathbf{L}(R)^{-1} = \left[\frac{w_{ij}(R, j)}{w(R)} \right]_{i,j \in S \setminus R}.$$

From these lemmas we obtain the following corollaries.

COROLLARY 1.1. *Let* $\mathbf{A} = (a_{ij})_{i,j=1}^{s-1}$ *be an* $(s-1) \times (s-1)$ *matrix and*

$$\mathbf{L} := \begin{pmatrix} \mathbf{A} & \mathbf{l} \\ 0 & \dots & 0 \end{pmatrix}, \quad where \quad \mathbf{l} = \left(-\sum_{j=1}^{s-1} a_{1j}, \dots, -\sum_{j=1}^{s-1} a_{s-1,j} \right)^T.$$

Moreover for $i, j = 1, \dots, s-1$ *and* $R \subseteq S$ *let* $F(s)$, $F_{ij}(s, j)$ *and* $F(S \setminus R)$ *be the relevant sets of forests in the graph* $g(\mathbf{L})$. *Then*:

(1) ([BoMa]) $\det \mathbf{A} = w(F(s))$;
(2) *if* $w(F(s)) \neq 0$, *then*

$$\mathbf{A}^{-1} = \left[\frac{w(F_{ij}(s, j))}{w(F(s))} \right]_{i,j=1}^{s-1}.$$

COROLLARY 1.2. *Let* $\mathbf{A} = (a_{ij})_{i,j=1}^{s-1}$, $\mathbf{x} = (x_1, \dots, x_{s-1})^T$, $\mathbf{b} = (b_1, \dots, b_{s-1})^T$, $\mathbf{x}, \mathbf{b} \in \mathbb{C}^{s-1}$. *Set*

$$\mathbf{L} := \begin{pmatrix} \mathbf{A} & \mathbf{l} \\ -\mathbf{b}^T & b \end{pmatrix}, \quad where$$

$$\mathbf{l} := \left(-\sum_{j=1}^{s-1} a_{1j}, \dots, -\sum_{j=1}^{s-1} a_{s-1,j} \right)^T \quad and \quad b := \sum_{j=1}^{s-1} b_j.$$

Moreover let $F(i)$, $i \in S$, *be the relevant sets of forests in the graph* $g(\mathbf{L})$. *Then*:

(1) *the system* $\mathbf{A}^T \mathbf{x} = \mathbf{b}$ *has exactly one solution if and only if* $w(F(s)) \neq 0$;
(2) *if* $w(F(s)) \neq 0$, *then* $x_i = w(F(i))/w(F(s))$.

1.3. *Directed forest expansions for characteristics of Markov chains.* Many characteristics of MCs are solutions of the systems of linear equations

(1.1)
$$\mathbf{L}(R)\mathbf{x} = \mathbf{b}$$

or

(1.2)
$$\mathbf{L}^T(R)\mathbf{x} = \mathbf{b},$$

where $\mathbf{b}$ is a nonnegative $(s - |R|)$-vector. From Lemma 1.2 or from Corollaries 1.1 and 1.2 we can easily obtain directed forest expansions for these characteristics. We give simple examples.

By a *stationary distribution* of an MC induced by a laplacian matrix $\mathbf{L}$ we mean a nonnegative, normalized vector $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_s)^T$ which is a solution of the system

(1.3)
$$\boldsymbol{\pi}^T \mathbf{L} = \mathbf{0}_s^T.$$

To solve (1.3) it is sufficient to solve the system

(1.4)
$$\mathbf{L}_{11}^T \mathbf{a}_1 = -(l_{12}, \ldots, l_{1s})^T,$$

and then to normalize the vector $\mathbf{a}^T := (1, \mathbf{a}_1^T)$. Obviously (1.4) is an example of (1.2).

A nonempty subset M of the state set S is called a *closed set* in the graph g if there are no states $i \in M$, $j \in S \setminus M$ so that $(i, j) \in E$. A *closed class* in g is, by definition, any closed set in g which is minimal for the order induced by inclusion. It is known that in every graph there is at least one closed class. It is clear that $w(f) > 0$ in the graph $g(\mathbf{L})$ if $\mathbf{L}$ is an MC laplacian matrix. Moreover if MC has one closed class then there is at least one state $i \in S$ such that $w(i) > 0$.

THEOREM 1.1 (Markov Chain Tree Theorem). *If an MC has one closed class, then*

$$\pi_i = \frac{w(i)}{\sum_{j \in S} w(j)} \quad \textit{for } i \in S.$$

The history of discovery of Theorem 1.1 remains mysterious. Aldous [Al] wrote that it is "the most often rediscovered result in probability theory". Kohler and Vollmerhaus [KoVo] called it the "diagram method" and attributed to Hill [Hi]. The Markov Chain Tree Theorem was proved independently by Freidlin and Wentzell [FreWe 1] and Shubert [Sh]. Our proof using the Matrix Tree Theorem seems to be new.

For $R \subseteq S$, $\omega \in \Omega$, $A \in \mathcal{F}$, $i, j \notin R$, $k \in R$ set:

- $\tau_R(\omega) := \inf\{t \geq 0 : X_t(\omega) \in R\}$, the hitting time of the set R,
- $\mathbf{Pr}_i(A) := \mathbf{Pr}(A \mid X_0 = i)$,
- $\mathbf{E}_i Y := \int_\Omega Y(\omega)\, \mathbf{Pr}_i(d\omega)$ for every measurable function $Y : \Omega \to \mathbb{R}$,
- $\mu_{ij}(R) = \mathbf{E}_i[\sum_{0 \leq t < \tau_R} \mathbf{1}(X_t = j)]$, the mean number of visits in j before absorption by R,

- $m_i(R) := \mathbf{E}_i \tau_R$, the mean hitting time of R,
- $p_{ik}(R) := \mathbf{Pr}_i\{X_{\tau_{R(\omega)}}(\omega) = k\}$, the probability distribution in the hitting time of R.

The last three characteristics may be computed by solving systems (1.1). For example, the vector $\mathbf{m}(R) = (m_i(R))_{i \in S \setminus R}$ is the solution of the system

$$\mathbf{L}(R)\mathbf{m}(R) = \mathbf{e}.$$

THEOREM 1.2. *Let $\mathbf{L}$ be an MC laplacian matrix such that there exists a forest with root R in the graph $g(\mathbf{L})$. Then for $i, j \in S \setminus R$ and $k \in R$,*

$$(1) \qquad\qquad \mu_{ij}(R) = \frac{w_{ij}(R, j)}{w(R)},$$

$$(2) \qquad\qquad p_{ik}(R) = \frac{w_{ik}(R)}{w(R)},$$

$$(3) \qquad\qquad m_i(R) = \frac{\sum_{j \notin R} w_{ij}(R, j)}{w(R)}.$$

Parts (2) and (3) of Theorem 1.2 were proved by Freidlin and Wentzell in the case of discrete time [FreWen 1–2]. Part (1) seems to be new.

Theorems 1.1 and 1.2 provide directed forest expansions for the most known characteristics of MCs. In the same way one can "expand" other parameters, e.g. the limiting matrix, the fundamental matrix or higher moments of the hitting time (see [Po 1]). The next application of directed forest expansions is in bounding eigenvalues of an MC laplacian matrix and, what is the same, of a transition probability matrix or a generator.

Let $0 = \lambda_1, \lambda_2, \ldots, \lambda_s$ be the eigenvalues of an MC laplacian matrix $\mathbf{L}$. Suppose that they are all real (this assumption is satisfied for the interesting reversible MCs—see Section 4) and numbered in increasing order: $\lambda_1 = 0 \leq \lambda_2 \leq \ldots \leq \lambda_s$. Set

$$F^k := \bigcup_{R \subseteq S, \, |R| = k} F(R) \quad \text{for } k = 0, \ldots, s.$$

Obviously $w(F^0) = 0$ and $w(F^s) = 1$.

THEOREM 1.3. *Let $\mathbf{L}$ be an MC laplacian matrix with one closed class which has only real eigenvalues. Then for $k = 2, \ldots, s$,*

$$\binom{s-1}{k-2}^{-1} \frac{w(F^{k-1})}{w(F^k)} \leq \lambda_k \leq \binom{s-1}{k-1} \frac{w(F^{k-1})}{w(F^k)}.$$

In the paper [Po2] there are related bounds between uncoupling measures and eigenvalues of a general MC Laplacian matrix.

2. Direct methods for systems of linear equations related to Markov chains. The directed forest expansions given in the previous sec-

tion can be used to bounding the roundoff error of direct methods which solve (1.1) or (1.2). This is interesting in some applications where accurate computations are needed, for example in models of transmission of high-definition television signals (see [HeRe], [O'C] for more details). To analyse the error of the algorithms we will consider perturbations of an MC laplacian matrix **L** and a nonnegative vector **b** caused by representation and computing in floating-point arithmetic with unit roundoff error ε.

For a given $k \in \mathbb{N}$, $0 < \varepsilon_1 < 1$, and functions $A, B : (0, \varepsilon_1) \to \mathbb{R}$, the notation $A(\varepsilon) = \langle k \rangle B(\varepsilon)$ means that

$$(1 - \varepsilon)^k \leq \frac{A(\varepsilon)}{B(\varepsilon)} \leq (1 - \varepsilon)^{-k} \quad \text{for } \varepsilon \in (0, \varepsilon_1)$$

(cf. Stewart [St], p. 407).

A family $\{\mathbf{L}(\varepsilon) : \varepsilon \in (0, \varepsilon_1)\}$ of MC laplacian $s \times s$ matrices is said to be a *relatively perturbed Markov chain* (RPMC) induced by an MC laplacian matrix $\mathbf{L} = (l_{ij})_{i,j \in S}$ if for every $i, j \in S$ with $i \neq j$,

$$-l_{ij}(\varepsilon) = \langle 1 \rangle l_{ij}.$$

A family $\{\mathbf{b}(\varepsilon) : \varepsilon \in (0, \varepsilon_1)\}$ of nonnegative u-vectors is called a *relatively perturbed nonnegative vector* (RPNV) induced by a nonnegative vector $\mathbf{b} = (b_i)_{i=1}^u$ if for every $i = 1, \ldots, u$,

$$b_i(\varepsilon) = \langle 1 \rangle b_i.$$

Note that $g(\mathbf{L}(\varepsilon)) = g(\mathbf{L})$ for $\varepsilon \in (0, \varepsilon_1)$. It is easy to prove the following proposition.

PROPOSITION 2.1. *Let* $\{\mathbf{L}(\varepsilon) : \varepsilon \in (0, \varepsilon_1)\}$ *be an RPMC induced by a laplacian matrix* $\mathbf{L}$. *Set* $u := s - |R|$. *Then*:

(1) $w(f)(\varepsilon) = \langle u \rangle w(f)$ *for* $f \in F(R)$,
(2) $w(R)(\varepsilon) = \langle u \rangle w(R)$,

where $w(f)(\varepsilon)$ *and* $w(R)(\varepsilon)$ *denote the weight of the forest* f *and the weight of* $F(R)$ *in* $g(\mathbf{L}(\varepsilon))$.

The next theorem says to what extent the entrywise relative error in **L** and **b** affects the error of solutions of (1.1) and (1.2).

THEOREM 2.1. *Let* $\{\mathbf{L}(\varepsilon) : \varepsilon \in (0, \varepsilon_1)\}$ *be an RPMC induced by* **L** *and* $R \subseteq S$. *Furthermore let* $\{\mathbf{b}(\varepsilon) : \varepsilon \in (0, \varepsilon_1)\}$ *be an RPNV induced by* **b** *of order* $u := s - |R|$. *Assume that there exists a forest with root* R *in* $g(\mathbf{L})$. *Then*:

(1) *the solutions* $\mathbf{x} = (x_i)_{i \in S \setminus R}$ *and* $\mathbf{x}(\varepsilon) = (x_i(\varepsilon))_{i \in S \setminus R}$ *of the systems*

$$\mathbf{L}(R)\mathbf{x} = \mathbf{b} \quad \text{and} \quad \mathbf{L}(R)(\varepsilon)\mathbf{x}(\varepsilon) = \mathbf{b}(\varepsilon)$$

satisfy

$$x_i(\varepsilon) = \langle 2u \rangle x_i \quad \text{for } i \in S \setminus R;$$

(2) *the solutions* $\mathbf{x} = (x_i)_{i \in S \setminus R}$ *and* $\mathbf{x}(\varepsilon) = (x_i(\varepsilon))_{i \in S \setminus R}$ *of the systems*

$$\mathbf{L}^T(R)\mathbf{x} = \mathbf{b} \quad \text{and} \quad \mathbf{L}^T(R)(\varepsilon)\mathbf{x}(\varepsilon) = \mathbf{b}(\varepsilon)$$

satisfy

$$x_i(\varepsilon) = \langle 2u \rangle x_i \quad \text{for } i \in S \setminus R.$$

The proof of the above theorem uses directed forest expansions given in Section 1. In the case of a stationary distribution we are led to the following corollary.

COROLLARY 2.1. *Let* $\{\mathbf{L}(\varepsilon) : \varepsilon \in (0, \varepsilon_1)\}$ *be an RPMC induced by a laplacian matrix* $\mathbf{L}$ *with one closed class. Then for* $i \in S$,

$$\pi_i(\varepsilon) = \langle 2(s-1) \rangle \pi_i,$$

where $\pi_i(\varepsilon)$ *are the components of the stationary distribution of an MC with laplacian matrix* $\mathbf{L}(\varepsilon)$.

O'Cinneide [O'C] obtained slightly weaker bounds without using directed forest expansions. His example indicates that they are nearly best possible.

It is known that subtractions appearing in computing a stationary distribution from the system (1.4) by Gaussian elimination can sometimes be the major source of roundoff errors. Grassmann, Taksar and Heyman [GrTaHe] introduced a procedure for this problem which involves no subtractions. This method is commonly referred to as the GTH algorithm. In [Po 1] we generalized the GTH algorithm for the systems (1.1) and (1.2). Theorem 2.2 below gives a bound on the entrywise relative error for these algorithms. The proof relies on Theorem 2.1.

Let x, y be floating point numbers with unit roundoff error ε. Moreover, let $\text{fl}(x \diamond y)$ for $\diamond \in \{+, -, *, /\}$, denote the result of the operation "$\diamond$" in floating-point arithmetic. Suppose that

$$\text{fl}(x \diamond y) = \langle 1 \rangle (x \diamond y)$$

and that arithmetic operations do not produce overflow or underflow.

THEOREM 2.2. *Let the assumptions of Theorem* 2.1 *hold. Then the vector* $\overline{\mathbf{x}}(\varepsilon) = (\overline{x}_i(\varepsilon))_{i \in S \setminus R}$ *computed by algorithm* 3.1 (3.2) *from* [Po 1] *satisfies the relation*

$$\overline{x}_i(\varepsilon) = \langle \psi(u) \rangle x_i, \quad \text{where} \quad \psi(u) = 5u^2 + 13u - 16.$$

If additionally $\psi(u)\varepsilon \leq 0.1$, *then*

$$|\overline{x}_i(\varepsilon) - x_i| \leq 1.06\psi(u)x_i\varepsilon \quad \text{for } i \in S \setminus R.$$

It is surprising that the above bounds do not depend on the condition numbers. Theorem 2.2 is a generalization and a sharpened version of

O'Cinneide's result [O'C, Th. 3], because it deals not only with (1.4) and there is no assumption that pivots are computed in double precision. In the same manner we can prove bounds for a given characteristic of MCs.

3. Aggregation algorithms for powerly perturbed Markov chains. Markov chains that appear in many applications (e.g. in queueing network analysis) are large and sparse. Their laplacian matrices have a nearly block structure. Such chains are referred to as *nearly completely decomposable MCs* (*NCDMCs*) and may be defined in the simplest case as a family $\{\mathbf{L}(\varepsilon) : \varepsilon \in (0, \varepsilon_1)\}$ of irreducible MCs indexed by a small parameter ε such that

$$\mathbf{L}(\varepsilon) = \begin{pmatrix} \mathbf{L}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{L}_m \end{pmatrix} + \varepsilon \mathbf{L}',$$

where $\mathbf{L}_1, \dots, \mathbf{L}_m$ are irreducible MC laplacian matrices of order $s_1, \dots, s_m$, respectively, and $\mathbf{L}'$ is an MC laplacian matrix of order $s = s_1 + \dots + s_m$.

For NCDMCs, direct methods can lead to immense fill-in during the triangularization part of computation. Furthermore NCDMCs have eigenvalues close to 1. This implies that standard iterative algorithms converge very slowly.

The idea of aggregation algorithms is to divide the problem into subproblems that can be solved nearly independently and then to link the subproblem solutions together (see Ch. 6 in [Ste-W] for more details).

NCDMCs have some generalizations. For example:

(1) The *linearly perturbed MCs*

$$\mathbf{L}(\varepsilon) = \mathbf{L}_0 + \varepsilon \mathbf{L}_1,$$

where $\mathbf{L}_0$ and $\mathbf{L}_1$ are MC laplacian matrices (see [HasHav], [Sch 1–3]).

(2) The *polynomially or analytically perturbed MCs*

$$\mathbf{L}(\varepsilon) = \sum_{n=0}^{N} \varepsilon^n \mathbf{L}_n,$$

where every $\mathbf{L}_n$ is an MC laplacian matrix, $N \leq \infty$ (see [HasHav], [RoWi 1–2]).

Below we define a wider class of perturbed MCs.

For given functions $A, B : \mathbb{R} \to \mathbb{R}$, the notation $A(\varepsilon) \sim B(\varepsilon)$ means that

$$\lim_{\varepsilon \to 0} \frac{A(\varepsilon)}{B(\varepsilon)} = 1.$$

To unify the notation we set $A(\varepsilon) \sim 0$ if there exists $\varepsilon_1 \neq 0$ such that for every $\varepsilon \in (-\varepsilon_1, \varepsilon_1)$, $A(\varepsilon) = 0$. Furthermore put $\frac{0}{0} := 1$.

A family $\{\mathbf{L}(\varepsilon) : \varepsilon \in (0, \varepsilon_1)\}$ of MC laplacian $s \times s$ matrices is said to be a *powerly perturbed Markov chain* (PPMC) if there exist matrices $\boldsymbol{\Delta} = (\delta_{ij})_{i,j \in S}$, $\mathbf{D} = (d_{ij})_{i,j \in S}$ with $\delta_{ij} \geq 0$ and $d_{ij} \in \mathbb{R} \cup \{\infty\}$ such that for every $i, j \in S$ with $i \neq j$,

$$-l_{ij}(\varepsilon) \sim \delta_{ij} \varepsilon^{d_{ij}}.$$

A family $\{\mathbf{b}(\varepsilon) : \varepsilon \in (0, \varepsilon_1)\}$ of nonnegative u-vectors is called a *powerly perturbed nonnegative vector* (PPNV) if there exist vectors $\boldsymbol{\zeta} = (\zeta_i)_{i=1}^u$ and $\mathbf{z} = (z_i)_{i=1}^u$ with $\zeta_i \geq 0$ and $z_i \in \mathbb{R} \cup \{\infty\}$ such that for every $i = 1, \ldots, u$,

$$b_i(\varepsilon) \sim \zeta_i \varepsilon^{z_i}.$$

From now on we identify a PPMC $\{\mathbf{L}(\varepsilon) : \varepsilon \in (0, \varepsilon_1)\}$ with the matrices $\boldsymbol{\Delta}$ and $\mathbf{D}$, and a PPNV $\{\mathbf{b}(\varepsilon) : \varepsilon \in (0, \varepsilon_1)\}$ with the vectors $\boldsymbol{\zeta}$ and $\mathbf{z}$.

Set

$$g^*(\mathbf{D}) := (S, \{(i, j) \in S \times S : d_{ij} < \infty\}).$$

Let f be a forest and F a set of forests in $g^*(\mathbf{D})$, respectively. Let us introduce some parameters of PPMCs:

$$d(f) := \sum_{(i,j) \in f} d_{ij}, \quad \delta(f) := \prod_{(i,j) \in f} \delta_{ij},$$

$$d(F) := \min_{f \in F} d(f), \quad \delta(F) := \sum_{f \in F : d(f) = d(F)} \delta(f).$$

Moreover, let $w(f)(\varepsilon)$ and $w(F)(\varepsilon)$ denote the weight of the forest f and the set F of forests in $g(L(\varepsilon))$.

It is easy to prove the following proposition.

PROPOSITION 3.1. *Let matrices $\boldsymbol{\Delta}$ and $\mathbf{D}$ be a PPMC. Furthermore let f and F be a forest and a set of forests in $g^*(\mathbf{D})$. Then:*

(1) $w(f)(\varepsilon) \sim \delta(f) \varepsilon^{d(f)}$;
(2) $w(F)(\varepsilon) \sim \delta(F) \varepsilon^{d(F)}$.

The following theorem describes the asymptotics of solutions of systems (1.1) and (1.2) connected to PPMCs in terms of directed forest expansions.

THEOREM 3.1. *Let matrices $\boldsymbol{\Delta}$ and $\mathbf{D}$ be a PPMC and $R \subseteq S$. Moreover let vectors $\boldsymbol{\zeta}$ and $\mathbf{z}$ of order $u := s - |R|$ be an PPNV. Suppose that there exists a forest with root R in $g^*(\mathbf{D})$. Then:*

(1) *the solution $\mathbf{x}(\varepsilon) = (x_i(\varepsilon))_{i \in S \setminus R}$ of the system*

$$\mathbf{L}(R)(\varepsilon)\mathbf{x}(\varepsilon) = \mathbf{b}(\varepsilon)$$

satisfies

$$x_i(\varepsilon) \sim \alpha_i \varepsilon^{a_i}, \quad \text{for } i \in S \setminus R;$$

(2) *the solution* $\mathbf{x}(\varepsilon) = (x_i(\varepsilon))_{i \in S \setminus R}$ *of the system*

$$\mathbf{L}^T(R)(\varepsilon)\mathbf{x}(\varepsilon) = \mathbf{b}(\varepsilon)$$

satisfies

$$x_i(\varepsilon) \sim \alpha'_i \varepsilon^{a'_i} \quad \text{for } i \in S \setminus R,$$

where the coefficients α_i, a_i, α'_i *and* a'_i *are some constants.*

In the proof of the above theorem (see [Po 1]) we use Lemma 1.2 and Proposition 3.1. The coefficients α_i, a_i, α'_i and a'_i are expressed by explicit but complicated formulas in terms of the parameters $d(F)$ and $\delta(F)$ for some sets of forests in $g^*(\mathbf{D})$ (see Th. 4.1 in [Po 1] for more details). In the simplest case of stationary distribution we have the following corollary.

COROLLARY 3.1. *Let* $\{\mathbf{L}(\varepsilon) : \varepsilon \in (0, \varepsilon_1)\}$ *be a PPMC induced by matrices* $\boldsymbol{\Delta}$ *and* $\mathbf{D}$ *such that the graph* $g^*(\mathbf{D})$ *has one closed class. Then*

$$\pi_i(\varepsilon) \sim \eta_i \varepsilon^{h_i} \quad \text{for } i \in S,$$

where

$$h_i := d(F(\{i\})) - \min_{j \in S} d(F(\{j\})),$$

$$\eta_i := \delta(F(\{i\})) / \sum_{j:h_j=0} \delta(F(\{j\})).$$

This corollary is similar to the results by Freidlin and Wentzell [FreWe 1–2] and Hwang and Sheu ([HwSh 1–2]), where a broader family of perturbed MCs is considered. However, the results there are less conclusive.

The formulas for α_i, a_i, α'_i and a'_i referred to above are not suitable for computation due to exponential complexity. In [Po 1] we give effective and accurate aggregation algorithms for these coefficients. They require $O(s^3)$ comparisons and arithmetic operations (algorithms 4.3–4.6 therein). In the proofs of correctness of the algorithms Theorem 4.1 is applied. Our algorithms are generalizations of methods by:

• Schweitzer [Sch 1–2], who constructs an algorithm for a stationary distribution of linearly perturbed, irreducible MCs;

• Hassin and Haviv [HasHav], who construct an algorithm for orders of magnitude of mean hitting times for linearly perturbed, irreducible MCs;

• Desai, Kumar and Kumar [DeKuKu], who construct an algorithm for orders of magnitude of a stationary distribution of special PPMCs.

4. Markov chain Monte Carlo algorithms for estimating integrals. In this section we consider the problem of approximation of the integral of a function $\mathbf{f} : S \to \mathbb{R}$ with respect to a probability distribution $\boldsymbol{\pi} = (\pi_i)_{i \in S}$, under the assumption that $\pi_i > 0$ for all $i \in S$. This problem

arises in statistical physics, for example when we estimate global characteristics of the Ising model. The state space S is very large (e.g. 2^{1000}) and direct summation is impossible. We use Monte Carlo algorithms which give estimates based on a relatively small sample drawn from S. The best general reference here is Sokal [So]. We are interested in the *Markov chain Monte Carlo* (MCMC) algorithms which generate discrete time MCs $X = (X_t)_{t \geq 0}$ with stationary distribution $\boldsymbol{\pi}$. The sample mean

$$\bar{f}_t(\omega) := \frac{1}{t} \sum_{j=0}^{t-1} f(X_j(\omega)), \quad \omega \in \Omega,$$

is the "natural" estimator of $\boldsymbol{\pi}^T \mathbf{f} := \sum_{i \in S} f_i \pi_i$. Here and in the sequel, f_i and $\mathbf{f}(i)$ have the same meaning. The ergodic theorem leads one to believe that, as the sample size t increases, the error of approximation becomes vanishingly small, because

$$\mathbf{Pr}\{\lim_{t \to \infty} \bar{f}_t = \boldsymbol{\pi}^T \mathbf{f}\} = 1.$$

To clarify the association with statistical mechanics, we write $\boldsymbol{\pi}$ in the form of "Gibbs distribution"

$$\pi_i(\tau) := \frac{\exp(-u_i/\tau)}{\sum_{j \in S} \exp(-u_j/\tau)}, \quad i \in S.$$

with a "potential" function $\mathbf{u} : S \to \mathbb{R}$ and with "temperature" τ. For simplicity of notation we write $\varepsilon := \exp(-1/\tau)$.

The best known of MCMC methods, the Metropolis algorithm [Me et al.], is the following. Let $\boldsymbol{\Delta} = (\delta_{ij})_{i,j \in S}$ be a symmetric irreducible stochastic matrix.

1) Let i be the state of the algorithm at time t, $X_t = i$. One chooses at random a "neighbour" $Y_t = j$ of i, according to a probability distribution $\delta_{i\cdot}$ given by the ith row of $\boldsymbol{\Delta}$.

2) If $u_j \leq u_i$, the state moves to j, $X_{t+1} := j$. Otherwise, the state moves to j with probability $\varepsilon^{u_j - u_i}$, or stays at i with probability $1 - \varepsilon^{u_j - u_i}$, $X_{t+1} := i$.

The Metropolis algorithm generates the MC with the following transitions:

$$\mathbf{Pr}^\varepsilon\{X_{t+1} = j \,|\, X_t = i\} = \delta_{ij}\varepsilon^{(u_j - u_i) \vee 0} \quad \text{for } j \neq i.$$

The Gibbs sampler demands a special structure of the state space. Let $S = K^L$, where L is a finite lattice and K is a finite set of "levels" with $|K| > 1$. For a "site" $x \in L$ and $i \in S$, let

$$N_x(i) = \{j \in S : j(z) = i(z) \text{ for all } z \neq x, \ z \in L\}$$

and $N(i) = \bigcup_{x \in L} N_x(i)$. For $i, j \in S$, put

$$g_x^\varepsilon(i, j) = \begin{cases} \varepsilon^{u_j}/(\sum_{k \in N_x(i)} \varepsilon^{u_k}) & \text{if } j \in N_x(i), \\ 0 & \text{otherwise.} \end{cases}$$

The transition probabilities of the Gibbs sampler (with random updating scheme) are the following:

$$\mathbf{Pr}^\varepsilon\{X_{t+1} = j \mid X_t = i\} = \frac{1}{|L|} \sum_{x \in L} g_x^\varepsilon(i, j).$$

In most applications we are interested in the behaviour of errors of algorithms as $\varepsilon \to 0$. To unify the analysis of MCMC methods, we introduce the following definition.

A family $\{\mathbf{L}(\varepsilon) : \varepsilon \in (0, \varepsilon_1)\}$ of MC laplacian $s \times s$ matrices is called Θ *powerly perturbed Markov chain* (ΘPPMC) if there exist numbers $c_0, c_1 > 0$ and a matrix $\mathbf{D} = (d_{ij})_{i,j \in S}$ with $d_{ij} \in \mathbb{R} \cup \{\infty\}$ such that for all $i, j \in S$ with $i \neq j$,

$$c_0 \varepsilon^{d_{ij}} \leq -l_{ij}(\varepsilon) \leq c_1 \varepsilon^{d_{ij}} \qquad (\varepsilon^\infty := 0).$$

Note that for the Metropolis algorithm and the Gibbs sampler we have, respectively,

$$d_{ij} := \begin{cases} (u_j - u_i) \vee 0 & \text{if } \delta_{ij} > 0, \\ \infty & \text{if } \delta_{ij} = 0; \end{cases}$$

$$d_{ij} := \begin{cases} u_j - \min_{k \in N_x(i)} f(k) & \text{if } j \in N_x(i) \text{ for some } x \in L, \\ \infty & \text{otherwise.} \end{cases}$$

It is clear that the family of ΘPPMC is larger than that of PPMC and has similar properties. For example,

$$(c_0/c_1)^{s-1} \varepsilon^{h_i} \leq \pi_i(\varepsilon) \leq (c_1/c_0)^{s-1} \varepsilon^{h_i}.$$

Let

$$v_k := d(F^{k-1}) - d(F^k), \quad \text{where} \quad F^k := \bigcup_{R \subseteq S, \, |R| = k} F(R)$$

and $d(F)$ is defined in Section 3.

Theorem 1.3 allows us to bound eigenvalues of ΘPPMCs.

THEOREM 4.1. *Let* $\{\mathbf{L}(\varepsilon) : \varepsilon \in (0, \varepsilon_1)\}$ *be a* Θ*PPMC which has only real eigenvalues, induced by a matrix* $\mathbf{D}$. *Then for* $k = 2, \ldots, s$,

$$\binom{s-1}{k-2}^{-1} \binom{s}{k}^{-1} k s^{s-k-1} c_0^{s-k+1} c_1^{k-s} \varepsilon^{v_k} \leq \lambda_k(\varepsilon)$$

and

$$\lambda_k(\varepsilon) \leq \binom{s-1}{k-1} \binom{s}{k-1} (k-1) s^{s-k} c_0^{k-s} c_1^{s-k+1} \varepsilon^{v_k}.$$

This theorem is similar to the result by Wentzell [We], where a larger family of chains than our ΘPPMC is considered. However, Wentzell's conclusion is less precise. In that paper, a fact equivalent to Lemma 1.1(1) is announced without proof. Chiang and Chow [ChiCho] proved that the coefficients v_k are the same for the Metropolis algorithm and the Gibbs sampler. Ingrassia [In] bounded $\lambda_2(\varepsilon)$ for these procedures using the Poincaré inequality ([Al], [DiSt], [Si]). In comparison with the Ingrassia inequality, Theorem 3.1 gives worse constants for λ_2, but allows us to bound all eigenvalues.

To establish bounds on errors of MCMC algorithms, we will use a recent result of Dinwoodie.

An MC with a laplacian matrix $\mathbf{L}$ is called *reversible* if it has a stationary distribution $\boldsymbol{\pi} = (\pi_i)_{i \in S}$ such that for all $i, j \in S$,

$$\pi_i l_{ij} = \pi_j l_{ji}.$$

Both the Metropolis algorithm and the Gibbs sampler generate reversible MCs. One can easily prove that the eigenvalues of reversible $\mathbf{L}$ are real. Let us number them in increasing order $\lambda_1 = 0 \leq \lambda_2 \leq \ldots \leq \lambda_s \leq 2$.

LEMMA 4.1 (Dinwoodie [Din]). *Let* $\mathbf{f} : S \to \mathbb{R}$. *Assume that* $0 \leq \mathbf{f} \leq 1$. *Then for every* $\delta \in [0, (8\lambda_2 + 16)^{-3}]$ *and* $i \in S$,

$$\mathbf{Pr}_i\{\bar{f}_t - \boldsymbol{\pi}^T \mathbf{f} \geq \delta\} \leq \left[1 + \frac{9\delta(\lambda_2 + 2)}{\sqrt{\pi_i}}\right] \exp(-t\lambda_2 \delta^2/2).$$

The following theorem yields bounds for errors of estimation of an integral by a sample mean for ΘPPMCs. The main advantages of this result are: large generality and explicit dependence on parameters of ΘPPMCs, ε and t.

THEOREM 4.2. *Let* $\{\mathbf{L}(\varepsilon) : \varepsilon \in (0, \varepsilon_1)\}$ *be a reversible and irreducible* $\Theta PPMC$ *induced by a matrix* $\mathbf{D}$ *and constants* c_0, c_1. *Let* $\boldsymbol{\pi}(\varepsilon)$ *be the stationary distribution of a MC laplacian matrix* $\mathbf{L}(\varepsilon)$. *Moreover, let* $\mathbf{f} : S \to \mathbb{R}$, $i \in S$, $p > 0$ *and* $t \in \mathbb{N}$. *Set*

$$r(\mathbf{f}) := \max_{i \in S} f_i - \min_{i \in S} f_i.$$

Then:

(1) *for* $\delta \in [0, r(\mathbf{f})(8C_1(c_0, c_1)\varepsilon^{v_2(\mathbf{D})} + 16)^{-3}]$,

$$\mathbf{Pr}_i^\varepsilon\{|\bar{f}_t^\varepsilon - \boldsymbol{\pi}^T(\varepsilon)\mathbf{f}| \geq \delta\}$$
$$\leq C_2((c_0/c_1)^{s-1}\varepsilon^{h_i(\mathbf{D})}) \exp[-tC_0(c_0, c_1)\varepsilon^{v_2(\mathbf{D})}\delta^2/(2r^2(\mathbf{f}))];$$

(2) $(\mathbf{E}_i^\varepsilon|\bar{f}_t^\varepsilon - \boldsymbol{\pi}^T(\varepsilon)\mathbf{f}|^p)^{1/p}$
$$\leq C_3(p, (c_0/c_1)^{s-1}\varepsilon^{h_i(\mathbf{D})}, r(\mathbf{f}))/\sqrt{tC_0(c_0, c_1)\varepsilon^{v_2(\mathbf{D})}},$$

where C_0, C_1, C_2, C_3 *are some constants*.

Theorem 4.2 and the results by Chiang and Chow [ChiCho] mentioned above support the empirical experience that the Metropolis algorithm and the Gibbs sampler have asymptotically equivalent behavior for low temperatures.

5. Markov chain Monte Carlo algorithms for finding a global minimum. In this section we investigate stochastic algorithms for searching a minimum of a function $\mathbf{f} : S \to \mathbb{R}$. These algorithms generate MCs and are applied when S is a large set (for example in the area of VLSI design). One of them, the Simulated Annealing (SA) algorithm, generates a nonhomogeneous MC with transition probabilities

$$\mathbf{Pr}\{X_{t+1} = j \mid X_t = i\} = \delta_{ij}\varepsilon_t^{(f_j - f_i)\vee 0} \quad \text{for } j \neq i,$$

where $(\varepsilon_t)_{t\geq 0}$ is a sequence decreasing to 0. We refer to [KiGeVe] and [RomSa] for a general exposition and applications of this method. It is easily seen that SA is a nonhomogeneous version of the Metropolis algorithm. Similarly we can modify the Gibbs sampler.

We are interested in asymptotic correctness of minimization algorithms, that is, in convergence of $\min_{0\leq s\leq t} f(X_s)$ to $\min_{i\in S} f(i)$ with probability one. To unify the analysis we introduce the following definitions.

Let $(X_t)_{t\geq 0}$ be a nonhomogeneous Markov chain with discrete time on a finite state space S. Suppose that for every $i \in S$, $\mathbf{Pr}\{X_0 = i\} > 0$. This condition is not particularly restrictive but it will allow us to omit some tedious details and concessions.

For simplicity of notation, write

$$\{A \text{ ult.}\} := \{\omega \in \Omega : \exists_{N\geq 0} \forall_{t\geq N} \ X_t(\omega) \in A\},$$
$$\{A \text{ i.o.}\} := \{\omega \in \Omega : \forall_{N\geq 0} \exists_{t\geq N} \ X_t(\omega) \in A\},$$
$$\{j \text{ i.o.}\} := \{\{j\} \text{ i.o.}\},$$

for all $A \subseteq S$ and $j \in A$.

A state $j \in S$ is called *recurrent* if $\mathbf{Pr}\{j \text{ i.o.}\} > 0$. Otherwise j is *transient*.

The *asymptotically closed class* (ACC) of a nonhomogeneous MC is, by definition, the subset R of S satisfying the following conditions:

(1) $\{R \text{ i.o.}\} \neq \emptyset$ a.s.;
(2) $\{R \text{ i.o.}\} = \{R \text{ ult.}\}$ a.s.;
(3) R is a set with properties (1) and (2) which is minimal with respect to inclusion.

The proposition below expresses the basic features of ACCs.

PROPOSITION 5.1. *Let* $R_1, \ldots, R_m$ *be all ACCs of a nonhomogeneous MC. Moreover, let* $T := S \setminus \bigcup_{i\leq m} R_i$. *Then*:

(1) $m \geq 1$;

(2) *the sets $R_1, \ldots, R_m, T$ are a partition of S*;

(3) $\{T \text{ i.o.}\} = \emptyset$ *a.s.*;

(4) $\{R_1 \text{ ult.}\} \cup \ldots \cup \{R_m \text{ ult.}\} = \Omega$ *a.s.*

An ACC R is a *recurrent class* if it satisfies additionally the following condition:

$$\{R \text{ i.o.}\} = \bigcap_{j \in R} \{j \text{ i.o.}\}.$$

In general, an ACC is not necessarily a recurrent class but for homogeneous MCs, both these notions reduce to the notion of a closed class (see Section 2.5 of [Io]).

A nonhomogeneous Markov chain $(X_t)_{t \geq 0}$ on the state space S is said to be a chain with *powerly diminishing transitions* (PDTC) if for all $i, j \in S$ with $i \neq j$,

$$c\varepsilon_t^{d_{ij}} \leq \mathbf{Pr}\{X_{t+1} = j \mid X_t = i\} \leq C\varepsilon_t^{d_{ij}},$$

where $C, c > 0$, $0 < \varepsilon_{t+1} \leq \varepsilon_t < 1$ for $t \geq 0$, $\lim_{t \to 0} \varepsilon_t = 0$, $0 \leq d_{ij} \leq \infty$, $\varepsilon_t^{\infty} := 0$.

The family of PDTC contains MCs generated by SA if we set, for $i \neq j$,

$$d_{ij} := \begin{cases} (u_j - u_i) \vee 0 & \text{if } \delta_{ij} > 0, \\ \infty & \text{otherwise.} \end{cases}$$

In the sequel we will consider only PDTCs.

The *recurrence order* of i, denoted by α_i, is defined to be the number

$$\alpha_i := \sup\left\{c \geq 0 : \mathbf{Pr}\left\{\omega \in \Omega : \sum_{t \geq 0} \varepsilon_t^c \mathbf{1}(X_t(\omega) = i) = \infty\right\} > 0\right\}, \quad i \in S$$

(we set $\sup \emptyset = -\infty$).

Note that a state $i \in S$ of a PDTC is recurrent if and only if $\alpha_i \geq 0$. Furthermore $\alpha_i \leq \varrho$, where $\varrho := \sup\{c \geq 0 : \sum_{t \geq 0} \varepsilon_t^c = \infty\}$. Suppose that

$$(5.1) \qquad\qquad\qquad\qquad \sum_{t \geq 0} \varepsilon_t^{\varrho} = \infty.$$

To state the main result of this section, Theorem 5.1, it will be convenient to modify definitions of a directed forest and its parameters, given in Sections 1 and 3.

A *forest on a domain $A \subseteq S$*, $A \neq \emptyset$, is a subgraph $f = (A, E_f)$ of $g(\mathbf{L})$ without cycles, in which from every state $i \in A$ there is at most one outgoing edge. Let $F_A(R)$ be the set of all forests in $g(\mathbf{L})$ on the domain A with root $R \subseteq A$. Similarly to the definitions given in Section 3, we introduce $d_A(f)$ and $d_A(F_A(R))$. For simplicity, we write $d_A(i)$ in place of $d_A(F_A(\{i\}))$ for $i \in A$. The analogues of the coefficients h_i in Corollary 3.1 are

$$h_A(i) := d_A(i) - \min_{j \in A} d_A(j).$$

For $\emptyset \neq A \subset S$, let
$$V(A) := \min_{i \in R,\, j \in S \setminus A} [h_A(i) + d_{ij}]$$
(we set $\min \emptyset := \infty$, $V(S) := \infty$ and $\infty - \infty = \infty$).

A *cup* in the graph $g^*(\mathbf{D})$ is a minimal set $A \subset S$ such that $V(A) \geq \varrho$.

It can be proved that in every graph $g^*(\mathbf{D})$ there is at least one cup and that two different cups in $g^*(\mathbf{D})$ are disjoint.

The main result of this section decribes recurrent classes and recurrent orders of PDTCs in terms of directed forest expansions.

THEOREM 5.1. (1) *For every $A \subseteq S$, A is a recurrent class if and only if A is a cup.*

(2) *For every $i \in S$, if i belongs to some cup, then $\alpha_i = \varrho - h_A(i)$; otherwise $\alpha_i = -\infty$.*

The results of Connors and Kumar and their method of solving "balance equations" for similar recurrence orders
$$\beta_i := \sup \left\{ c \geq 0 : \sum_{t \geq 0} \varepsilon_t^c \mathbf{Pr}(X_t = i) = \infty \right\}$$
were a starting point of the paper [NiPo]. In this paper the tail σ-field of a PDTC was characterized in terms of the recurrence orders
$$\gamma_i(\omega) := \sup \left\{ c \geq 0 : \sum_{t \geq 0} \varepsilon_t^c \mathbf{1}(X_t(\omega) = i) = \infty \right\}.$$

The γ_is and cups were expressed there by balance equations, without using directed forest expansions. Moreover, we proved that the solutions of the balance equations are unique. Borkar [Bo], using a similar technique, derived balance equations for γ_i. Niemiro [Ni] applied the description of the tail σ-field to analyse convergence in probability for PDTCs generated by SA algorithms. It is worth noting that, unlike β_i, the recurrence orders α_i are uniquely determined and can be efficiently computed.

If assumption (5.1) is not satisfied, to prove Theorem 5.1 it is sufficient to replace the condition "$V(A) \geq \varrho$" by "$V(A) > \varrho$" in the definition of a cup.

The theorem leads to an explicit criterion of reachability (a.s.) of every set A for PDTCs.

COROLLARY 5.1. *For a PDTC and every $A \subseteq S$, the following conditions are equivalent*:

(1) $\mathbf{Pr}\{A \ i.o.\} = 1$.
(2) $\mathbf{Pr}\{\exists_{t \geq 0} X_t \in A\} = 1$.
(3) *In every cup there is a state which belongs to A.*

Application of this corollary to MCs generated by the SA algorithm and to the set $S^* := \{i \in S : \forall_{j \in S} \ u_i \le u_j\}$ of global minima yields Connors and Kumar's theorem [ConKu].

References

[Al]   D. Aldous, *Reversible Markov chains and random walks on graphs*, preprint, 1994.

[Bo]   V. S. Borkar, *Pathwise recurrence orders and simulated annealing*, J. Appl. Probab. 29 (1992), 472–476.

[BoMa]   R. Bott and J. P. Mayberry, *Matrices and trees*, in: *Economic Activity Analysis*, O. Morgenstern (ed.), Wiley, New York, and Chapman & Hall, London, 1953, 391–400.

[Cha]   S. Chaiken, *A combinatorial proof of the all minors matrix tree theorem*, SIAM J. Algebraic Discrete Methods 3 (1982), 319–329.

[Che]   W.-K. Chen, *Applied Graph Theory*, *Graphs and Electrical Networks*, 2nd ed., North–Holland, New York, 1976.

[ChiCho] T.-S. Chiang and Y. Chow, *Asymptotic behavior of eigenvalues and random updating schemes*, Appl. Math. Optim. 28 (1993), 259–275.

[ConKu] D. P. Connors and P. R. Kumar, *Simulated annealing type Markov chains and their balance equations*, SIAM J. Control Optim. 27 (1989), 1440–1461.

[CvDoSa] D. M. Cvetković, M. Doob and H. Sachs, *Spectra of Graphs—Theory and Applications*, Deutscher Verlag Wiss., Berlin, 1979, and Academic Press, New York, 1979.

[DeKuKu] M. Desai, S. Kumar and P. R. Kumar, *Quasi-statically cooled Markov chains*, Probab. Engrg. Inform. Sci. 8 (1994), 1–19.

[DiSt]   P. Diaconis and D. Stroock, *Geometric bounds for eigenvalues of Markov chains*, Ann. Appl. Probab. 1 (1991), 36–61.

[Din]   I. H. Dinwoodie, *A probability inequality for the occupation measure of a reversible Markov chain*, ibid. 5 (1995), 37–43.

[FieSe] M. Fiedler and J. Sedlácek, *O w-basich orientovaných grafu*, Čas. Pest. Mat. 83 (1958), 214–225 (in Czech).

[FreWe 1] M. I. Freidlin and A. D. Wentzell, *On small random perturbations of dynamical systems*, Russian Math. Surveys 25 (1970), 1–55.

[FreWe 2] —, —, *Random Perturbations of Dynamical Systems*, Springer, New York, 1984.

[GrTaHe] W. K. Grassmann, M. I. Taksar and D. P. Heyman, *Regenerative analysis and steady-state distributions for Markov chains*, Oper. Res. 33 (1985) 1107–1116.

[HasHav] R. Hassin and M. Haviv, *Mean passage times and nearly uncoupled Markov chains*, SIAM J. Discrete Math. 5 (1992), 386–397.

[HeRe]   D. P. Heyman and A. Reeves, *Numerical solution of linear equations arising in Markov chain model*, ORSA J. Comput. 1 (1989), 52–60.

[Hi]   T. L. Hill, *Studies in irreversible thermodynamics IV. Diagrammatic representation of steady state fluxes for unimolecular systems*, J. Theoret. Biol. 10 (1966), 442–459.

[HwSh 1]   C.-R. H w a n g and S.-J. S h e u, *Large-time behavior of perturbed diffusion Markov processes with applications to the second eigenvalue problem for Fokker–Planck operators and simulated annealing*, Acta Appl. Math. 19 (1990), 253–295.

[HwSh 2]   —, —, *Singular perturbed Markov chains and exact behaviors of simulated annealing processes*, J. Theor. Probab. 5 (1992), 223–249.

[In]   S. I n g r a s s i a, *On the rate of convergence of the Metropolis algorithm and Gibbs sampler by geometric bounds*, Ann. Appl. Probab. 4 (1994), 347–389.

[Io]   M. I o s i f e s c u, *Finite Markov Processes and Their Applications*, Wiley, 1980.

[KeSn]   J. G. K e m e n y and J. L. S n e l l, *Finite Markov Chains*, Van Nostrand, Princeton, 1960.

[KiGeVe]   S. K i r k p a t r i c k, C. D. G e l a t t and M. P. V e c c h i, *Optimization by simulated annealing*, Science 220 (1983), 671–680.

[KoVo]   H. H. K o h l e r and E. V o l l m e r h a u s, *The frequency of cyclic processes in biological multistate systems*, J. Math. Biol. 9 (1980), 275–290.

[Me et al.]   W. M e t r o p o l i s, A. R o s e n b l u t h, M. R o s e n b l u t h, A. T e l l e r and E. T e l l e r, *Equations of state calculations by fast computing machines*, J. Chem. Phys. 21 (1953), 1087–1092.

[Mo]   B. M o h a r, *The Laplacian spectrum of graphs*, in: Y. Alavi *et al.* (eds.), *Graph Theory*, *Combinatorics and Applications*, Wiley, New York, 1991, 871–898.

[Ni]   W. N i e m i r o, *Limit distributions of Simulated Annealing Markov chains*, Discussiones Math. 15 (1993), 241–269.

[NiPo]   W. N i e m i r o and P. P o k a r o w s k i, *Tail events of some nonhomogeneous Markov chains*, Ann. Appl. Probab. 5 (1995), 261–293.

[O'C]   C. A. O'C i n n e i d e, *Entrywise perturbation theory and error analysis for Markov chains*, Numer. Math. 65 (1993), 109–120.

[Po 1]   P. P o k a r o w s k i, *Directed forests and algorithms related to Markov chains*, Inst. Math., Polish Acad. Sci., 1997 (in Polish).

[Po 2]   —, *Uncoupling measures and eigenvalues of stochastic matrices*, J. Appl. Anal. 4 (1998), 261–269.

[RoWi 1]   J. R. R o h l i c e k and A. S. W i l l s k y, *The reduction of Markov generators*: *An algorithm exposing the role of transient states*, J. Assoc. Comput. Mach. 35 (1988), 675–696.

[RoWi 2]   —, —, *Multiple time scale decomposition of discrete time Markov chains*, Systems Control Lett. 11 (1988), 309–314.

[RomSa]   F. R o m e o and A. S a n g i o v a n n i-V i n c e n t e l l i, *A theoretical framework for simulated annealing*, Algorithmica 6 (1991), 367–418.

[Sch 1]   P. J. S c h w e i t z e r, *Perturbation theory and finite Markov chains*, J. Appl. Probab. 5 (1968), 401–413.

[Sch 2]   —, *Perturbation series expansions of nearly completely decomposable Markov chains*, in: J. W. Cohen, O. J. Boxma and H. C. Tijm (eds.), *Telegraphic Analysis and Computer Performance Evaluation*, Elsevier, North-Holland, Amsterdam, 1986.

[Sh]   B. O. S h u b e r t, *A flow-graph formula for the stationary distribution of a Markov chain*, IEEE Trans. Systems Man. Cybernet. 5 (1975), 565–566.

[Si]   A. S i n c l a i r, *Improved bounds for mixing rates of Markov chains and multicommodity flow*, Combin. Probab. Comput. 1 (1992), 351–370.

[So]   A. D. S o k a l, *Monte Carlo Methods in Statistical Mechanics*: *Foundations and New Algorithms*, Cours de Troisième Cycle de la Physique en Suisse Romande, Lausanne, June 1989 (unpublished).

[Ste]   G. W. S t e w a r t, *Introduction to Matrix Computations*, Academic Press, New York, 1973.

[Ste-W]   W. J. S t e w a r t, *Introduction to the Numerical Solution Markov Chains*, Princeton Univ. Press, Princeton, 1994.

[We]   A. D. W e n t z e l l, *On the asymptotics of eigenvalues of matrices with elements of order* $\exp(V_{ij}/2\varepsilon^2)$, Dokl. Akad. Nauk SSSR 202 (1972), 263–265 (in Russian); English transl.: Soviet Math. Dokl. 13 (1972), 65–68.

Piotr Pokarowski   Current address:
Institute of Mathematics   Institute of Applied Mathematics and Mechanics
Polish Academy of Sciences   Warsaw University
Śniadeckich 8   Banacha 2
00-950 Warsaw, Poland   02-097 Warszawa, Poland
E-mail: pokar@impan.gov.pl   E-mail: pokar@hydra.mimuw.edu.pl