Perfect Simulation and Backward Coupling

S.G. Foss† and R.L. Tweedie‡
Institute of Mathematics, Novosibirsk and Colorado State University

Abstract

Algorithms for perfect or exact simulation of random samples from the invariant measure of a Markov chain have received considerable recent attention following the introduction of the “coupling-from-the-past” (CFTP) technique of Propp and Wilson. Here we place such algorithms in the context of backward coupling of stochastically recursive sequences. We show that although general backward couplings can be constructed for chains with finite mean forward coupling times, and can even be thought of as extending the classical “Loynes schemes” from queueing theory, successful CFTP algorithms can be constructed if and only if the chain is uniformly geometric ergodic. We also relate the convergence moments for backward coupling methods to those of forward coupling times: the former typically lose at most one moment compared to the latter.

---

*Work supported in part by NSF Grant DMS-9504561 and by CRDF Grant RM1-226
†Postal Address: Institute of Mathematics, 630090, Novosibirsk, Russia; email: foss@math.nsc.ru
‡Postal Address: Department of Statistics, Colorado State University, Fort Collins CO 80523, USA; email: tweedie@stat.colostate.edu

Keywords: Irreducible Markov chains, invariant measures, stationary measures, coupling from the past, stochastic recursive sequences, uniform ergodicity, backward coupling, stochastic monotonicity

AMS Subject classification: 60J10, 60K05, 60K30
1 Introduction

One of the most compelling contributions of Marcel Neuts to applied probability is the emphasis he placed on computable methods. As just one example, his introduction of models based on phase-type distributions, and the ability they give to practitioners to actually calculate the invariant measures of Markov chains [15], has led to a real revolution in the use of computationally feasible probabilistic thinking.

This paper is concerned with a different approach to calculation of invariant measures. Nonetheless, its emphasis is on the development and properties of an algorithmic approach, and one that is becoming widely used since its recent introduction in [17]. Although our pathwise constructions lead to a different style of algorithm than that introduced by Marcel Neuts, they are very much in the spirit of his thinking, and it is with great pleasure that we contribute to this volume dedicated to him.

Although most of his work was on countable space Markov chains, one of us remembers fondly a conversation with Marcel in Manila that led to a general space formulation of his matrix geometric methods [22]. This paper similarly gives results on general spaces, although many of them will find their implementation most feasible on finite or countable spaces.

We consider a Markov chain on a state space $X$, which we assume is a separable metric space, although many of our results hold without this assumption: it is needed however to ensure that for any two random variables $X, Y$ the set $\{X = Y\}$ is measurable. This restriction could be relaxed as noted on p.18 of [2] but we do not take that step here. Since we are concerned with sample path behaviour, we write $X = \{X_n\}_{n=0}^{\infty}$ for the version of the chain starting at $x_0$ (which may be fixed or random); versions starting with other initial values will be distinguished by other notation when they occur.

Let $P(x, A)$ denote the transition law of the chain, and, when it exists, denote its unique invariant (or stationary) measure by $\pi$. By stationarity, $\pi$ satisfies

$$\pi(A) = \int_X \pi(dx) P(x, A)$$

(1)

for all measurable $A$. Let $P^n(x, A)$ denote the $n$-step transition law of the chain. Frequently, under appropriate irreducibility and aperiodicity assumptions [12], $P^n(x, \cdot)$ converges to $\pi$ in the sense that

$$\lim_{n \to \infty} ||P^n(x, \cdot) - \pi|| = 0$$

(2)

for each $x \in X$; here, $|| \cdot ||$ denotes the total variation norm. Our coupling methods will typically lead to convergence results such as (2).

In this paper we will describe algorithms that will enable the simulation of $\pi$ exactly (that is, by drawing a random sample known to be from $\pi$).
Of course, if the explicit form of $\pi$ is known, say by solving (1) directly, then there are many ways of drawing such a sample; however, except for models such as the matrix-geometric class described by Neuts [15], there are few cases where such analytic forms are tractable.

Without analytic knowledge of $\pi$, perhaps the most popular method of sampling at least approximately from $\pi$ is the obvious one: simulate the initial value $x_0$ of the chain from the given initial distribution and use $P$ to generate $X_n$ for successive values of $n$. An approximate sample from $\pi$ can be taken as the value of the chain after a suitably chosen (large) number of iterations provided (2) holds. How close the sample is to $\pi$ in a distributional sense depends upon the convergence rate of the chain to stationarity. Unfortunately, no general guidelines are available for deciding on an appropriate number of iterations for a general state space Markov model, although much work has been done on this problem recently in many different contexts [13, 18, 4].

In this paper we consider instead methods of “perfect” simulation, based on a backward coupling technique that enables exact draws from $\pi$. These are closely linked to the “coupling-from-the-past” (CFTP) algorithm introduced recently by Propp and Wilson [17] who showed it to be extremely effective in areas such as statistical physics; they give a description of related “perfect” methods in [24]. Following this, CFTP algorithms have recently been applied in spatial point processes [9] and Markov chain Monte Carlo areas [14], and many other applications appear likely in the near future.

As shown in [17], CFTP methods are particularly powerful in implementation when the chain possesses monotonicity properties. Many chains of interest in storage, network and queueing theory satisfy various stochastic monotonicity properties [6, 19, 20], and Wilson and Lund [23] examine the application of “coupling-from-the-past” in various of these models.

Our goals in this paper are twofold. Firstly, we show in Section 2 and Section 3 how perfect simulation can be viewed in the context of stochastic recursive sequences, a tool introduced in [1, 5, 3], and we show that these methods can be regarded as descendents of the now-classical “Loynes scheme” [11] for describing the invariant measure of a simple queuing model.

We derive convergence properties of these more general backward coupling algorithms in Section 6. In particular we show how the rates of convergence of backward coupling algorithms can be connected to those of forward coupling times of the chain, and that typically we lose at most one moment in going from forward to backward couplings.

Secondly, we show in Section 4 that CFTP itself can be viewed as a special case of backward coupling; and perhaps most surprisingly, we find that CFTP can be implemented successfully if and only if the Markov chain is uniformly ergodic: that is, the convergence in (2) is uniform (and then geometric) in the starting point.

If CFTP is not successful, then one may not be able to decide if one has reached a backward coupling time. However, in such circumstances we may be able to construct
verifiable backward coupling times which lead to estimates of \( \pi \) with known degrees of approximation: in [8] we show that this can be done for “horizontal” couplings from the past, as opposed to the “vertical” backward couplings associated with CFTP in this paper.

2 Stochastic Recursive Sequences

We now describe a theoretical framework that will lead to forward and backward coupling constructions on continuous spaces. We will use the “stochastic recursive sequence” (SRS) construction of \( \boldsymbol{X} \) developed by Borovkov and Foss [1, 5, 3], who considered general stability properties of these systems in both Markovian and non-Markovian cases. The SRS construction enables us to use deterministic sample path arguments which are particularly suited to a simulation environment, and to the study of stationary measures on which we focus here.

The SRS approach is based on the fact that [10, 2] that one can construct a probability space \((\Omega, \mathcal{F}, P)\), an independent and identically distributed sequence \(\{\xi_n\}_{n=-\infty}^{\infty}\) of (uniform) \(U[0,1]\) random variables, and a measurable function \(f : [0, \infty) \times [0,1] \to [0, \infty)\) such that \(\boldsymbol{X}\) satisfies the recursion

\[
X_0 = x_0, \quad X_{n+1} = f(X_n, \xi_n), \quad n \geq 0,
\]

and has transition probabilities \(P(x, \cdot)\). Note that at this stage we only use \(\xi_n, n \geq 0\); in the next section the doubly-infinite construction will become relevant.

There are infinitely many such constructions to choose from, depending on, for example, the joint distributions selected for \(f(x, \xi_1)\) and \(f(y, \xi_1)\) for different \(x\) and \(y\). The particular construction used is largely a matter of convenience.

**Example 1: Queueing models based on random walks** Perhaps the simplest and best-known versions of (3) occur in queueing theory. Let \(W_n\) be an i.i.d. sequence of random variables with distribution \(G\) on \(\mathbb{R}\), and define

\[
X_{n+1} = [X_n + W_{n+1}]^+, \quad n = 0, 1, 2, \ldots
\]

This random walk on \([0, \infty)\) occurs as a model in many areas of operations research [12]. By using the inverse probability transform on the distribution of \(W_n\) we can write each \(W_n = G^{-1}(\xi_n)\) where the \(\xi_n\) are i.i.d. \(U[0,1]\) variables; and then we have (3) holding with

\[
f(x, y) = [x + G^{-1}(y)]^+.
\]

**Example 2: Uniformly minorized chains** A more complex version of this construction is the following, which will prove central in Section 4. Suppose the chain satisfies the uniform minorization condition

\[
P(x, \cdot) \geq \beta \varphi(\cdot), \quad x \in \mathcal{X}
\]
for some probability measure $\varphi$ and some $0 < \beta \leq 1$. Then [12, Chapter 16] the chain is uniformly ergodic: that is, convergence in (2) is uniform in $x$ and indeed

$$\lim_{n \to \infty} ||P^n(x, \cdot) - \pi|| \leq [1 - \beta]^n, \quad x \in \mathcal{X}$$  \hfill (5)

One way to simulate from this chain is to draw a sequence $U_n$ of i.i.d. uniform random variables, and a sequence $U_n$ of i.i.d. variables with law $\varphi$; if $U_{n+1} \leq \beta$ then we set $X_{n+1} = V_{n+1}$, and if $U_{n+1} > \beta$ then we choose $X_{n+1}$ from the distribution $Q(X_n, \cdot) = [P(X_n, \cdot) - \beta \varphi(\cdot)]/[1 - \beta]$. It is obvious that $X_{n+1}$ has the correct marginal distribution from this construction.

Here again one can construct $\xi_n$ and $f$ as functions of $U_n$ and the inverse distribution functions of $\varphi$ and $P(x, \cdot)$. Although this is more complicated, it is clear how the joint distributions will interact: in particular, if $U_{n+1} \leq \beta$ it follows that the value of $X_{n+1}$ is independent of $X_n$. This is the "multigamma" coupler described in Murdoch and Green [14], who also describe related and more efficient forms of coupling in this context.

The attraction of the SRS representation (3) is that we go all the way to an i.i.d. sequence $\xi_n$, placing all of the structure in $f$, rather than using a (perhaps more obvious) generating scheme such as that in Example 2. This simplifies the analysis very considerably, and in particular it enables analytic approaches on each path via $f$ while keeping the random aspects of the chain in the simplest possible form.

Usually, it will be convenient to work with the canonical probability space $(\Omega, \mathcal{F}, P)$, where $\Omega = [0,1]^\mathbb{Z}$, $\mathcal{F}$ is the cylinder $\sigma$-algebra of Borel sets, and $P$ is Lebesgue measure. In this case, for $\omega = \{\omega_n\}_{n=-\infty}^{\infty} \in \Omega$, we can define the random variables using the simple coordinate map by $\xi_n(\omega) = \omega_n$. The structure of $X$ is then entirely determined by the function $f$; this is essentially found by the usual inversion of the distribution function of the transition law from $x$, as in Example 1.

Working on the canonical probability space, one can define both forward and backward couplings for $X$. We first describe the better known ideas of forward coupling, which we shall use in the sequel. It is worth noting that the ideas below work in general when $\{\xi_n\}$ form a stationary ergodic sequence, and not just an i.i.d. sequence: then $\{X_n\}$ can be analysed within the SRS framework [2] and forward and backward coupling are valuable tools for studying stationarity.

**Forward coupling.**

Let $X = \{X_n\}_{n=0}^{\infty}$ and $X' = \{X'_n\}_{n=0}^{\infty}$ be two Markov chains that differ in initial values only; specifically, we start the two chains with $X_0 = x_0$ and $X'_0 = x'_0$ (where $x_0$ and $x'_0$ are allowed to be random). The two chains then evolve via the recursions $X_{n+1} = f(X_n, \xi_n)$ and $X'_{n+1} = f(X'_n, \xi_n)$ for $n \geq 0$, and (critically) with the same $\{\xi_n\}_{n=0}^{\infty}$ driving both $X$ and $X'$.
Then a random integer valued time $\tau$ is called a forward coupling time if

$$\{\tau \leq n\} \Rightarrow \{X_{n+m} = X'_{n+m} \quad \forall m \geq 0\}. \quad (6)$$

If $\tau$ is a forward coupling time and $\tau'$ is any other integer-valued random variable such that $\tau' \geq \tau$ almost surely, then $\tau'$ is also a forward coupling time.

Given the recursive construction above, if $X_n = X'_n$, then $X_{n+m} = X'_{n+m}$ for all $m \geq 1$. Therefore it is possible to define the minimal forward coupling time

$$\tau(X, X') = \min\{n \geq 0 : X_n = X'_n\} \leq \infty \quad (7)$$

Forward coupling of $X$ and $X'$ is called successful if $\tau(X, X') < \infty$ almost surely.

More generally, let $\{X^{(t)}\}_{t \in T}$ be a family of Markov chains, with different initial values $X_0^{(t)}$. Then

$$\tau(\{X^{(t)}\}_{t \in T}) = \sup_{t_1, t_2 \in T} \tau(X^{(t_1)}, X^{(t_2)}) \quad (8)$$

is the minimal forward coupling time of $\{X^{(t)}\}_{t \in T}$, and any $\tau \geq \tau(\{X^{(t)}\}_{t \in T})$ is a forward coupling time of $\{X^{(t)}\}_{t \in T}$.

Forward coupling has become the classical method of proving convergence theorems such as (2). Suppose that $\pi$ is invariant for $P$, and define a stationary version $\overline{X}$ of the chain by choosing $\overline{X}_0$ from $\pi$. We will write $\overline{\tau}$ for the minimal forward coupling time $\tau(X, \overline{X})$. Then from the definition (6) we have immediately the coupling inequality

$$\|P^n(x_0, \cdot) - \pi\| \leq P(\overline{\tau} > n). \quad (9)$$

Thus convergence to $\pi$ in total variation is obtained if $\overline{\tau}$ is successful; and rates of convergence in (2) follow if we know moments of the coupling time $\overline{\tau}$.

It is well known (cf. [12, Chapter 13]) that for an irreducible chain on the integers (or indeed more generally for $\pi$-irreducible chains), the coupling $\overline{\tau}$ is successful from $\pi$-almost all $x_0$ whenever a stationary measure exists.

As a specific example we can consider the coupling in Example 2, based on (4). If we use this construction, then $\overline{\tau} \leq \tau_U$ where $\tau_U = \min\{n \geq 0 : U_n \leq \beta\}$, and

$$X_{\tau_U} = V_{\tau_U}, \quad (10)$$

not just in distribution but almost surely. Thus using (9), since

$$P(\tau_U > n) = (1 - \beta)^n \quad (11)$$

we recover the geometric convergence of the uniformly ergodic chains in (5).
3 Backward Coupling

Forward coupling is a valuable tool to investigate existence and rate of convergence results, but clearly it does not allow us to find variables that have the actual distribution \( \pi \): note, for example, that in Example 2, from (10) we have \( X_{\tau_{\mu}} \sim \varphi \), so that at the forward coupling time \( \tau_{\mu} \), the chain will definitely not be in the stationary regime.

We now introduce the related concept of backward coupling, which addresses this problem. To define the backward coupling construction, we move to a doubly infinite time-axis.

We main tool we need to introduce is the shift transformation \( \theta \). On the canonical probability space \((\Omega, \mathcal{F}, P)\), let \( \theta^m \) denote the \( m \)-shift transformation; specifically, for \( \omega = \{\omega_n\}_{n=-\infty}^{\infty} \in \Omega \) and \( m \in \mathbb{Z} \), set
\[
\theta^m(\omega) = \{\omega_{n+m}\}_{n=-\infty}^{\infty}.
\]

(12)

Clearly, \( \theta^{m+n}(\omega) = \theta^m(\theta^n(\omega)) \) for all \( m, n \in \mathbb{Z} \) and \( \omega \in \Omega \). It also follows that \( \xi_{n+m}(\omega) = \xi_n(\theta^m(\omega)) \) for any \( m, n \in \mathbb{Z} \) and \( \omega \in \Omega \). For a set \( B \in \mathcal{F} \), define \( \theta^m B = \{\theta^m \omega : \omega \in B\} \).

For any random variable \( \psi : \Omega \to X \), the shifted random variable \( \psi_m \) is now defined as \( \psi_m(\omega) = \psi(\theta^m(\omega)) \). It is crucial to note that, for any \( m \), the shift \( \theta^m \) is measure-preserving and that \( \psi \) and \( \psi_m \) coincide in law.

On the basis of the SRS construction, for any \( m \), we have thus defined a “shifted” chain \( \theta^m X = \{\theta^m X_n\}_{n=0}^{\infty} \). This has the following interpretation: the “original” Markov chain \( X \) starts from \( x_0 \) at time 0 and takes on the value \( X_n \) at time \( n \); the “shifted” chain \( \theta^m X \) starts at time \( m \) from \( \theta^m x_0 \) (which coincides with \( x_0 \) if \( x_0 \) is constant but not necessarily otherwise), and takes the value \( \theta^m X_n \) at time \( m + n, n = 1, 2, \ldots \)

Alternatively one can view \( \theta X_n \) as a construction of the shifted sequence of the \( \xi_i \). Since \( X_n \) is measurable with respect to the \( \sigma \)-algebra generated by the \( \xi_i, i < n \), it can be represented as a measurable (non-random) function of \( \xi_0, \ldots, \xi_{n-1} \)
\[
X_n = g(\xi_0, \ldots, \xi_{n-1});
\]

and therefore it follows that
\[
\theta X_n = g(\xi_1, \ldots, \xi_n).
\]

(13)

It is vital to note that since \( \theta^m \) is measure-preserving, \( X \) and \( \theta^m X \) coincide in law.

**Backward coupling**

The random variable \( \nu \leq \infty \) is a **backward coupling time** for \( X \) if
\[
\{\nu \leq m\} \Rightarrow \{m \geq 0 : \theta^{-n_1} X_{n_1} = \theta^{-n_2} X_{n_2} \ \forall n_1, n_2 \geq m\}.
\]
A backward coupling time is *successful* if $\nu < \infty$ almost surely.

Note that, unlike a forward coupling time, the backward coupling time just depends on the chain starting from $x_0$. We will see in the next section how this relates to coupling of different copies of the chain from different starting points.

As with forward coupling, we can define the *minimal backward coupling time* $\nu(X)$ by

$$\nu(X) = \min\{m \geq 0 : \theta^{-n_1}X_{n_1} = \theta^{-n_2}X_{n_2} \quad \forall n_1, n_2 \geq m\} \leq \infty$$ (14)

If $\nu$ is a backward coupling time and $\nu'$ is any integer-valued random variable such that $\nu' \geq \nu$ almost surely then $\nu'$ is also a backward coupling time.

If one has a successful backward coupling time, then one can give a constructive approach showing that there exists a stationary version of the chain $X$.

**Theorem 3.1** Let $\nu$ be a successful backward coupling time. Put $\tilde{X}^0 = \theta^{-\nu}X_\nu$ and define $\tilde{X}^n = \theta^n\tilde{X}^0$ for $n \in \mathbb{Z}$. Then the sequence $\tilde{X} = \{\tilde{X}^n\}_{n=-\infty}^{\infty}$ forms a stationary Markov chain with transition probabilities $P(x, \cdot)$, and for any $n$, $\tilde{X}$ satisfies the recursion

$$\tilde{X}^{n+1} = f(\tilde{X}^n, \xi_n).$$ (15)

for each $n \in \mathbb{Z}$ almost surely.

**Proof** The fact that the sequence $\tilde{X}^n$ is stationary follows since $\theta$ is measure preserving. The harder part is to show that $\tilde{X}^n$ is a Markov chain with the appropriate transition law. This will follow if we can prove that (15) holds, since any sequence of random variables satisfying the SRS construction using $f$ and the i.i.d. sequence $\xi_n$ must have these properties.

It is sufficient to prove (15) for $n = 0$ only: the general case is by induction. Manipulations give

$$\tilde{X}^1 = \theta\tilde{X}^0 = \theta(\lim_{m \to \infty} \theta^{-m}X_m) = \lim_{m \to \infty} \theta^{-m+1}X_m = \lim_{n \to \infty} \theta^{-n}X_{n+1} = \lim_{n \to \infty} \theta^{-n}f(X_n, \xi_n) = \lim_{n \to \infty} f(\theta^{-n}X_n, \xi_0) = f(\theta^{-\nu}X_\nu, \xi_0) = f(\tilde{X}^0, \xi_0);$$

in the last line we have relied on the fact that $\theta^{-\nu}X_\nu = \theta^{-n}X_n$ for all sufficiently large $n$ since $\nu$ is finite because the backward coupling is successful.

It is of considerable interest to note that there is a classical construction in queueing models, in the context of the SRS described in Example 1, where the structure of $\pi$ arises naturally from such a backwards coupling construction.
As far as we are aware, Loynes [11] was the first to introduce the backward shift for the stability study of multi-server queues and backward coupling for single-server queues. To describe the “Loynes scheme”, consider, for instance, Example 1 and assume $E W_1 < 0$ and $X_0 = 0$. Put

$$\nu = \min\{m \geq 0 : W_{-1} + \ldots + W_{-m} \leq 0 \ \forall n \geq m\}.$$ 

Note that the SLLN implies $\nu < \infty$ a.s. But now, since

$$X_n = \max(0, W_n, W_{n-1} + W_n, \ldots, W_1 + \ldots + W_n),$$

we have that

$$\theta^{-n} X_n = \theta^{-\nu} X_\nu$$

a.s. for all $n \geq \nu$ and, therefore, $\nu$ is a successful backward coupling time. Thus, from Theorem 3.1, we have a very explicit representation of $\pi$.

Theorem 3.1 proves in general that the existence of a successful backward coupling time guarantees the existence of (at least one) stationary measure for the chain. In general, the converse is not true. Example 3 below shows a positive recurrent chains where the backward coupling construction fails.

However, it does follow that under weak conditions, the backward coupling construction $\bar{X}^0 = \theta^{-\nu} X_\nu$ does give the structure of the invariant measure, thus showing that this is an appropriate way to attempt to construct invariant measures, as we now indicate.

Given the existence of a stationary measure, one can first show (cf. [10]) that there exists a doubly-infinite stationary sequence $\bar{X} = \{\bar{X}_n\}_{n=-\infty}^\infty$ satisfying $\bar{X}_{n+1} = f(\bar{X}_n, \xi_n)$ a.s. for all $n \in \mathbb{Z}$. Furthermore, $\bar{X}_n$ is measurable with respect to the $\sigma$-field generated by $\xi_{n-1}, \xi_{n-2}, \xi_{n-3}, \ldots$.

We now have from Theorem 4 of Borovkov and Foss [2], the following converse of Theorem 3.1, in the case where the forward coupling time between $X$ and the stationary chain $\bar{X}$ has finite mean.

**Theorem 3.2** Suppose there exists a stationary measure for $X$ and let $\bar{X}_{n+1} = f(\bar{X}_n, \xi_n), n \in \mathbb{Z}$ be some stationary version of the chain. Assume that, for some constant $X_0 = x_0$, the forward coupling time

$$\bar{\tau} = \min\{n \geq 0 : X_n = \bar{X}_n\} < \infty \ a.s. \quad (16)$$

Then $\bar{X}_{n+1} = \theta \bar{X}_n$ a.s., and if $E[\bar{\tau}] < \infty$, then there is a successful backward coupling time $\nu$ and $\theta^{-\nu} X_\nu = \bar{X}_0$ a.s. Hence for all $n$ we have a.s. $\bar{X}_n = \bar{X}^n$ where the representation $\bar{X}^n$ is defined in Theorem 3.1.
Conditions for which \( \mathbb{E}[\tilde{\tau}] < \infty \) are well-known: in [16] these are called chains recurrent of degree 2, and in [12, Chapter 13.4] their convergence properties are discussed. Equivalent conditions for a chain to have this property include:

(a) for any set \( B \) of positive \( \pi \)-measure, we have \( \int_B \pi(dx) \mathbb{E}_x[\tau_B^2] < \infty \);

(b) for any atom \( \alpha \) of positive \( \pi \)-measure, we have \( \mathbb{E}_\alpha[\tau_\alpha^2] < \infty \)

where \( \tau_B \) is the first hitting time on \( B \). Other conditions in terms of drift functions which imply ergodicity of degree 2 can be deduced from the results of [21].

Finally, we note that unlike forward coupling times, backward coupling times must be finite with probability one (i.e. successful) or infinite with probability one. To see this, note from (14) that we have

\[
\theta \nu = \min\{m \geq 0 : \theta^{-n_1+1}X_{n_1} = \theta^{-n_2+1}X_{n_2}, \ \forall n_1, n_2 \geq m\}
\leq \min\{m \geq 1 : \theta^{-n_1+1}f(X_{n_1-1}, \xi_{n_1-1}) = \theta^{-n_2+1}f(X_{n_2-1}, \xi_{n_2-1}), \ \forall n_1, n_2 \geq m\}
= \min\{m \geq 1 : f(\theta^{-n_1+1}X_{n_1-1}, \xi_0) = f(\theta^{-n_2+1}X_{n_2-1}, \xi_0), \ \forall n_1, n_2 \geq m\}
\leq \min\{m \geq 1 : \theta^{-n_1+1}X_{n_1-1} = \theta^{-n_2+1}X_{n_2-1} \ \forall n_1, n_2 \geq m\}
= 1 + \min\{l \geq 0 : \theta^{-n_1}X_{n_1} = \theta^{-n_2}X_{n_2} \ \forall n_1, n_2 \geq l\} = 1 + \nu.
\]

Therefore if \( \theta \nu = \infty \) we must have \( \nu = \infty \), and so \( \{\nu = \infty\} \) is invariant. Since the i.i.d. sequence \( \{\xi_n\} \) is ergodic, it follows that \( \mathbb{P}\{\nu = \infty\} \) is zero or one as stated.

4 Vertical backward coupling and CFTP

Theorem 3.1 is more of an existence result than an algorithm. It shows that, if we assume the chain admits a unique invariant measure \( \pi \), then if we could identify a successful backward coupling time \( \nu \), by following the shifted path forward from \( \nu \) to time zero we would get a perfect sample from \( \pi \).

This is similar to the goal of the Propp and Wilson [17] “coupling-from-the-past” algorithm. One of the key issues addressed by Propp and Wilson [17] is that of “verifiability”: when running their CFTP algorithm, which we describe below, one can in fact tell whether one is starting from a time point from which the coupling is guaranteed. In contrast, we note that the general definition (13) of a backward coupling time appears to require knowledge of the behaviour of the chain in the infinite past.
However, by constructing what we will call a \textit{vertical backward coupling time}, the Propp-Wilson algorithm does actually ensure that a given time can be verified as a successful backward coupling time in our terms.

The idea here is an elegant one. Suppose that we consider the complete family of chains \( X^{(z)} \equiv \{X_n^{(z)}\}_{n \geq 0} \) starting from every \( x \in X \). If we can find a time \( T \) such that \textit{all} of the shifted chains \( \theta^{-T} X^{(z)} \) starting at time \(-T\) have the same value \( \theta^{-T} X_T^{(z)} \) at time zero, then (as we show formally in Theorem 4.1) it follows that \( T \) is a backward coupling time; and thus from Theorem 3.1 this value is a perfect draw from \( \pi \).

Intuitively, it is clear why the backward coupling theorems work with this choice of \( T \). For consider a chain starting at \(-\infty\) with the stationary distribution \( \pi \). At every iteration it maintains the distribution \( \pi \); and from then on it follows the trajectory from that value. But of course it arrives at the same place at time zero no matter where it starts: so the value returned by the algorithm at time zero must itself be a draw from \( \pi \).

This is the content of Theorem 1 of Propp and Wilson [17], who give it for an irreducible aperiodic finite state space chain. They coin the term “coupling-from-the-past” (CFTP) for this algorithm, since in essence \(-T\) is a coupling time with the stationary version started at \(-\infty\). Murdoch and Green [14] discuss versions of the CFTP algorithm for some special classes of chains on continuous spaces, to which we shall return later.

We now show formally that \( T \) is a backward coupling time, thus putting this algorithm within the framework of the SRS theory in the previous section. It is worth noting as we do this that the version of the CFTP algorithm in Theorem 2 of [17] is virtually phrased in terms of a SRS framework, as is the structure of the continuous space versions in [14].

\textbf{Theorem 4.1} For any \( z, y \in X \), put

\[
T(z, y) = \min\{n \geq 0 : \theta^{-n} X_n^{(z)} = \theta^{-n} X_n^{(y)}\} \tag{17}
\]

and assume that

\[
T \equiv \sup_{z, y} T(z, y) \leq \infty. \tag{18}
\]

is a well defined random variable. Then \( T \) is a backward coupling time for the Markov chain that starts from any arbitrary \( x_0 \in X \); and so if \( T \) is successful, for any \( x_0 \) we have \( \theta^{-T} X_T^{(x_0)} \sim \pi \) where \( \pi \) is a unique invariant measure for the chain.

\textbf{Proof} Fix \( x_0 \in X \) and denote by \( X = \{X_n\} \) as usual the Markov chain that starts from \( x_0 \). Take any \( y \in X \) and set \( \overline{X^0} = \theta^{-T} X_T^{(y)} \), where \( T \) is defined by (17).
Now for any \( m \geq T \), and for any \( z \in X \), if \( \theta^{-m}X_{(m-T)} = z \), then \( \theta^{-m}X_m = \theta^{-T}X^{(z)}_T \); but since \( T \geq T(z,y) \), we have \( \theta^{-T}X^{(z)}_T = \overline{X}^0 \).

Therefore, \( T \) is a backward coupling time for \( X \) as required. From Theorem 3.1 we have that there is then a stationary measure \( \pi \) such that \( \overline{X}^0 \sim \pi \); and if \( \pi' \) is any other invariant measure, then by drawing \( x_0 \) from \( \pi' \) and repeating the above proof we find that \( \pi' = \pi \) also.

It is not always true that \( T \) is a well-defined random variable, although it is valid under some natural assumptions such as continuity of the function \( f \). To consider the completely general case one can proceed rather as in the comments after Theorem 1 of [2], by defining an outer measure \( P^* \) as \( P^*(A) = \inf \{ P(B), A \subset B, B \in \mathcal{F} \} \), replacing the condition \( P(T < \infty) = 1 \) by \( P^*(T < \infty) = 1 \), and repeating the proof using \( P^* \).

We call \( T \) a vertical backward coupling time since intuitively we consider the states in \( X \) stacked vertically and then draw sample paths from each one to see if we have achieved a backward coupling.

Perhaps surprisingly, we can characterise precisely the class of chains for which \( T \) is successful, and for which we can implement CFTP.

**Theorem 4.2** There exists a successful vertical backward coupling time \( T \) for \( X \) if and only if \( P \) is uniformly ergodic.

**Proof** We first note that (as indicated in Lemma 5 of [17]) \( T \) is subadditive in the sense that

\[
P[T > m + n] \leq P[T > m]P[T > n]
\]

for any \( m, n \in \mathbb{Z}_+ \). This follows since if either of the events

\[
A_m = \{ \theta^{-m}X^{(x)}_m = \theta^{m}X^{(y)}_m \ \forall \ x, y \in X \}
\]

\[
A_{n,m} = \{ \theta^{-(m+n)}X^{(x)}_n = \theta^{-(m+n)}X^{(y)}_n \ \forall \ x, y \in X \}
\]

holds then

\[
A_{m+n} = \{ \theta^{-(m+n)}X^{(x)}_{m+n} = \theta^{-(m+n)}X^{(y)}_{m+n} \ \forall \ x, y \in X \}
\]

holds also: that is, if the paths either reach a constant at time zero from \( -m \), or reach a constant at time \( -(m+n) \) from \( -(m+n) \), then from any time before \( -(m+n) \) they must also reach a constant at time zero. But \( A_{n,m}, A_m \) are independent by construction of the \( \xi_i \), and by the measure preservation of \( \theta \), we also have \( A_{n,m} \) and \( A_n \) are identical in distribution: thus (19) holds.

Let us suppose then that \( T \) is successful. Since \( T < \infty \) a.s. then from (19) we get

\[
P[T > n] \leq c\lambda^n
\]

(20)
for some $c < \infty$ and some $\lambda < 1$.

Now we relate the distribution of $T$ to that of $\tau_x$, the forward coupling time for the family $\{X^{(e)}\}_{e \in \mathcal{E}}$. Again using the fact that $\theta$ is measure preserving, we have
\[
P(T > n) = P(\exists x, y : \theta^{-n}X^{(e)}_n \neq \theta^{-n}X^{(y)}_n)
\]
\[
= P(\exists x, y : X^{(z)}_n \neq X^{(y)}_n)
\]
\[
= P(\tau_x > n).
\]

Hence $\tau_x$ also has geometric tails from (20). But now we can use the (forward) coupling inequality (9), noting that $\tilde{\tau} \leq \tau_x$ a.s., to give for every $x_0$
\[
\|P^n(x_0, \cdot) - \pi\| \leq P(\tilde{\tau}) > n) \leq c\lambda^n;
\]
and so the chain is uniformly ergodic as claimed.

The proof that there is a successful vertical backward coupling for any uniformly ergodic chain is effectively the multigamma coupling construction of [14], outlined in Example 2. From Theorem 7.0.2 of [12] we know that for a uniformly ergodic chain there exists some measure $\varphi$, and some $m \geq 1$ and $0 < \beta \leq 1$ such that
\[
\text{that is, the } m\text{-skeleton chain } X_{mn} \text{ with transition law } P^m \text{ is uniformly minorized. Draw and fix an i.i.d. sequences } U_n \sim U[0,1] \text{ and } V_n \sim \varphi, \text{ for } n = \ldots, -1, 0, 1, \ldots.
\]
Now construct each step in the sample path of $X_{mn}$ by setting $X_{mn} = V_{mn}$ if $U_{nm} \leq \beta$ and otherwise drawing $X_{mn} \sim [P(X_{(n-1)m}, \cdot) - \beta\varphi(\cdot)]/\{1 - \beta\}$.

Now set $T_U = \min\{n \geq 1 : U_{nm} \leq \beta\}$. Then for all $n \geq T_U$, and for any particular $x_0$, we will have $\theta^{-n}X_{n-T_U} = V_{T_U}$ and so $T_U$ is a vertical backward coupling time as required.

This construction shows that for uniformly ergodic chains, the geometric tails guaranteed by (20) are actually at rate $\lambda \leq [1 - \beta]^{1/m}$ where $\beta$ is given by (23), and $c = 1$.

Murdoch and Green [14] show in examples that $T_U$ is not always an effective coupling in a practical sense, and they give some useful constructions to ensure that for chains with more structure (and in particular with "larger" local minorizing measures) one can reach a vertical backward coupling time more quickly in practice.

In Section 6 we give other results on the speed of convergence of backward coupling times, but we first show how one of the other ingredients of the Propp-Wilson approach, namely stochastic monotonicity, fits into the SRS framework.
5 Stochastically monotone chains

We let \( \geq \) denote a partial ordering on \( X \) and call the transition law \( P \) stochastically monotone if, writing \( X_n = f(X_{n-1}, \xi_n) \) as in (3), we have \( f \) monotone in its first coordinate: that is, \( f(x_1, y) \leq f(x_2, y) \) whenever \( x_1 \leq x_2 \).

For a chain on \( \mathbb{R}_+ \) that is stochastically monotone in the standard sense that \( P(x_1, [y, \infty)) \leq P(x_2, [y, \infty)) \) for \( x_1 \leq x_2 \), one can always choose \( f \) in this way [2].

This definition of monotonicity enables us to prove simply

**Theorem 5.1** Suppose the state space of the Markov chain is a partially ordered space with a finite set \( U \) of maximal and a finite set \( L \) of minimal elements, and that \( f \) is monotone in its first coordinate.

Let \( X^{(l)} \) be the chain starting from the minimal state \( x_l \in L \), and \( X^{(u)} \) be the chain starting from the maximal state \( x_u \in U \).

If we define

\[
T(l, u) = \min\{n \geq 0 : \theta^{-n} X^{(l)}_n = \theta^{-n} X^{(l)}_n\}
\]

and

\[
T = \max_{l \in L, u \in U} T(l, u)
\]

then \( T \) is a vertical backward coupling time for the Markov chain starting from any \( x \in X \).

Moreover, \( T \) has the same distribution as the forward coupling time \( \tau_{U, L} \) of the family of Markov chains \( \{X^{(l)}_n, X^{(u)}_n\}_{l \in L, u \in U} \).

**Proof** Since there are only a finite number of extremal elements of \( X \), for any \( x \) we can find \( l, u \) such that \( x_l \leq x \leq x_u \). Monotonicity of \( f \) implies that for any \( m \geq n \),

\[
\theta^{-n} X^{(l)}_n \leq \theta^{-m} X^{(l)}_m \leq \theta^{-m} X^{(z)}_m \leq \theta^{-m} X^{(u)}_m \leq \theta^{-n} X^{(u)}_n \quad \text{a.s.}
\]

Since \( \theta^{-T} X^{(l)}_T = \theta^{-T} X^{(u)}_T \),

\[
\theta^{-m} X^{(l)}_m = \theta^{-m} X^{(z)}_m = \theta^{-m} X^{(u)}_m
\]

for any \( m \geq T \). Therefore, \( T \) is a backward coupling time for the whole family \( \{X^{(z)}_n\}_{n \geq 0} \); i.e. it is a vertical backward coupling time as required.

Again we have by preservation of measure

\[
P(T > n) = P(\exists u, l : \theta^{-n} X^{(u)}_n \neq \theta^{-n} X^{(l)}_n)
\]

\[
= P(\exists u, l : X^{(u)}_n \neq X^{(l)}_n)
\]

\[
= P(\tau_{U, L} > n).
\]
where $\tau_{U,L}$ is the forward coupling time of the family $\{X^{(i)}, X^{(u)}\}_{i \in L, u \in U}$.  

It is intuitively clear why this class of stochastically monotone chains will be uniformly ergodic: for example the hitting times on any connected set can be expected to be sandwiched between those from extremal elements, and so the supremum of these hitting times will be finite for positive recurrent chains, which is another characterization of uniform ergodicity [12, Theorem 16.0.2].

If no minimal or maximal elements exist then in some cases one can "truncate" the state space and then derive approximations of the accuracy of the (perfectly sampled) truncated measure compared to the original measure. Details of such an approach are in [7].

6 Moments of backward coupling times

We finally consider in more detail the distribution of a general backward coupling time $\nu$, and find conditions under which it is successful; and also conditions under which we can guarantee rapid convergence of the backward coupling.

To do this we need the notion of a strong forward coupling time.

For any $m \geq 0$, introduce the Markov chain $^mX = \{^mX_n\}_{n \geq 0}$, where $^mX_n = \theta^{-m}X_{n+m}$. This Markov chain starts at time 0 from the random value $^mX_0 = \theta^{-m}X_m$. Then $\tau^0$ is called a strong coupling time for $X$ if $\tau^0$ is a coupling time for the family $^mX_{m \geq 0}$.

We then have the following link between the (minimal) strong forward coupling time $\tau^0$ and the (minimal) backward coupling time $\nu$.

**Theorem 6.1** The random variables $\nu$ and $\tau^0$ have identical distributions.

**Proof** For any $k \geq 0$, we have

\[
P(\tau^0 \leq k) = P(^mX_k = ^mX_k \quad \forall m_1, m_2 \geq 0)
\]

\[
= P(\theta^{-m_1}X_{m_1+k} = \theta^{-m_2}X_{m_2+k} \quad \forall m_1, m_2 \geq 0)
\]

\[
= P(\theta^{-k}\{\theta^{-m_1}X_{m_1+k} = \theta^{-m_2}X_{m_2+k} \quad \forall m_1, m_2 \geq 0\})
\]

\[
= P(\theta^{-m_1-k}X_{m_1+k} = \theta^{-m_2-k}X_{m_2+k} \quad \forall m_1, m_2 \geq 0)
\]

\[
= P(\theta^{-n_1}X_{n_1} = \theta^{-n_2}X_{n_2} \quad \forall n_1, n_2 \geq k)
\]

\[
= P(\nu \leq k).
\]
Thus, properties of $\nu$ can be translated into those of $\tau^0$. We note again however that $X_{\tau^0}$ and $X_\nu$ are not the same: hence, we cannot convert all results from backward couplings to results for forward couplings.

The strong coupling time $\tau^0$ (or equivalently $\nu$) can be expressed in terms of a supremum of ordinary coupling times as is now shown.

**Theorem 6.2** Let $\vec{\tau}$ be a coupling time of $X$ and a stationary chain $\{\vec{X}^n\}$ as in (16), and assume $E[\vec{\tau}] < \infty$. If we set $\vec{\tau}_m = \theta^m \vec{\tau}$, then

$$\tau^0 = \sup_{m \geq 0} (\vec{\tau}_m - m).$$

**Proof** Using Theorem 3.2 we see that $\nu < \infty$ almost surely so $\vec{X}^k = \theta^{-\nu} X_{\nu+k} = \lim_m \theta^{-m} X_{m+k}$. Thus,

$$\{\tau^0 \leq k\} = \{\vec{X}^k = mX_k \quad \forall m \geq 0\}$$

$$= \{\theta^{-m}\vec{\tau} = m \leq k \quad \forall m \geq 0\}$$

$$= \{\sup_{m \geq 0} (\vec{\tau}_m - m) \leq k\}.$$

Since $\tau^0 \geq \vec{\tau}$ a.s., $\tau^0$ cannot have more finite polynomial moments than $\vec{\tau}$. The following results show that $\tau^0$ can lose no more than one polynomial moment in comparison with $\vec{\tau}$. Furthermore, when $\vec{\tau}$ has an exponential tail, a strict upper bound for the tail of $\tau^0$ (and hence from Theorem 6.1 for the tail of $\nu$) can be given.

**Theorem 6.3** Let $\vec{\tau}$ be as in (16). Then

(i) For any $k > 1$, if $E[\vec{\tau}^k] < \infty$, then $E[(\tau^0)^{k-1}] < \infty$; and so $\nu$ loses at most one polynomial moment from $\vec{\tau}$.

(ii) If $E[\exp(A\vec{\tau})] \leq B$ for some $A, B > 0$, then

$$E[\exp(A\tau^0)] \leq Be^A/(e^A - 1).$$

and so $\nu$ has the same order of geometric convergence as $\vec{\tau}$.
\textbf{Proof} \quad We have from Theorem 6.2 and the measure preserving property of the shift that
\begin{align*}
\mathbb{E}[(\tau^0)^{k-1}] &= \sum_{l=1}^{\infty} l^{k-1} P(\tau^0 = l) \\
&\leq (k - 1) \sum_{l=1}^{\infty} \sum_{r=1}^{l} r^{k-2} P(\tau^0 = l) \\
&= (k - 1) \sum_{r=1}^{\infty} r^{k-2} P(\tau^0 \geq r) \\
&\leq (k - 1) \sum_{r=1}^{\infty} r^{k-2} \mathbb{E}[(\bar{r} - r - 1)^+] \\
&\leq (k - 1) \sum_{m=1}^{\infty} P(\bar{r} = m) \sum_{r=1}^{m} r^{k-2} (m - r) \\
&\leq k^{-1} \sum_{m=1}^{\infty} P(\bar{r} = m) (m + 1)^k \\
&\leq k^{-1} \mathbb{E} \{ \bar{r} + 1 \}^k \leq k^{-1} 2^k \mathbb{E} \{ \bar{r} \}^k.
\end{align*}
Thus (i) is established.

To prove (ii), note that for any $A > 0$ we have
\begin{align*}
\mathbb{E} \exp \{ A \tau^0 \} &= \left[ \sum_{n=1}^{\infty} P(\tau^0 \geq n) \exp \{ An \} \right] (e^A - 1) e^{-A} + 1 \\
&\leq \left[ \sum_{m=1}^{\infty} \sum_{n=m}^{\infty} P(\bar{r} \geq m) \exp \{ An \} \right] (e^A - 1) e^{-A} + 1 \\
&= \sum_{m=1}^{\infty} P(\bar{r} \geq m) e^{Am} + 1 \\
&= (\mathbb{E} [\exp \{ A \bar{r} \}] - 1) (e^A - 1)^{-1} e^A + 1 \\
&\leq B (e^A - 1)^{-1}.
\end{align*}
\hfill \Box

Example 3 below shows that one can lose exactly one moment as in (i). Conversely, one may lose no moments at all. For following Lemma 1 of [2], consider a stochastically monotone model with $f$ monotone in its first coordinate, as in Section 5, and suppose $x_0$ is the minimal element of the space. Then from monotonicity
\begin{align*}
X_0 \leq \theta^{-1} X_1 \leq \ldots \leq \theta^{-n} X_n \leq \ldots \theta^{-\nu} X^\nu \quad \text{a.s.}
\end{align*}
and so
\begin{align*}
P(\bar{r} \leq n) &= P(X_n = \bar{X}^n) = P(\theta^{-n} X_n = \bar{X}^0) \\
&= P(\theta^{-m} X_m = \bar{X}^0 \forall m \geq n) = P(\nu \leq n)
\end{align*}  \tag{29}
Conditions for \( \tilde{\tau} \) to have polynomial moments can be deduced from those given in [21]. For example, for the random walk on \([0, \infty)\) in Example 1 we see that \( \tilde{\tau} \) typically has polynomial moments of one order less than the tails of the increment distribution \( G \). Now (29) implies we need only the mean of \( G \) to exist in this case for both the backward and the forward couplings to succeed.

Conditions for \( \tilde{\tau} \) to have geometric moments as in (ii), or for the chain to be geometrically ergodic, are given in [12, Chapter 15]. Again, for the random walk in Example 1 we find that \( \tilde{\tau} \) typically has geometric moments provided the tails of the increment distribution \( G \) are exponential also. As pointed out in [12], this random walk is not uniformly ergodic, so in this case we cannot use a vertical backward coupling time approach. We show in [8] that this type of example can be approached at least approximately by a "horizontal" backward coupling scheme.

We conclude with an example to show that the strong coupling time may lose exactly one moment from the coupling time \( \tilde{\tau} \); and moreover that the strong coupling time, and hence the backward coupling time \( \nu \), may not be proper even if the Markov chain is itself positive recurrent. Thus, it may not be possible to use backward coupling to construct the stationary measure as in Theorem 3.1 unless \( E[\tilde{\tau}] < \infty \).

**Example 3** Suppose \( X = \{0, 1, 2, \ldots\} \), and take \( X_0 = 0 \) and \( \{\xi_n\} \) to be a sequence of i.i.d. positive integer-valued random variables. Put \( f(0, y) = y, f(1, y) = 1 \) and \( f(x, y) = x - 1 \) for \( x \geq 2 \). In this simple case the stationary chain is degenerate at \( \{1\} \), and so it is clear that \( \tilde{\tau} = \xi_0 \), and that

\[
\tau^0 = \min\{n \geq 0 : \xi_{l - 1} \leq l \quad \forall l \geq n\}.
\]

Thus for sufficiently large \( n \), by independence of the \( \xi_n \),

\[
P(\tau^0 > n) = P(\bigcup_{i=n}^{\infty} \{\xi_{l - 1} > l\})
\]

\[
= 1 - \exp\{\sum_{i=n}^{\infty} \log P(\xi_0 \leq l)\}
\]

\[
\approx 1 - \exp\{-\sum_{i=n}^{\infty} P(\xi_0 > l)\}. \tag{30}
\]

Hence if \( E[\xi_0] = \infty \), then \( P(\tau^0 > n) = 1 \) for all \( n \), and the backward coupling is a.s. infinite also; while if \( E[\xi_0] \) is finite, then \( \sum_{i=n}^{\infty} P(\xi_0 \geq l) \to 0 \) as \( n \to \infty \) and from (30) we have

\[
P(\tau^0 > n) \approx \sum_{i=n}^{\infty} P(\xi_0 > l).
\]

Therefore, \( \tilde{\tau} \) has a finite \( k \)-th moment if and only if \( \tau^0 \) has a finite \((k - 1)\)-st moment.

\[\square\]

Note that, in this example, 0 is a non-essential state and the chain is not irreducible although it is \( \delta \)-irreducible. One can construct more complicated examples of irreducible Markov chains with similar properties, but we omit the details.
Acknowledgements

We are grateful to David Wilson and Robert Lund for valuable discussions.

References


