Path Coupling: a Technique for Proving Rapid Mixing in Markov Chains

Russ Bubley Martin Dyer
School of Computer Studies
University of Leeds
Leeds LS2 9JT
United Kingdom

Abstract

The main technique used in algorithm design for approximating #P-hard counting problems is the Markov chain Monte Carlo method. *At the heart of the method is the study of the convergence (*mixing*) rates of particular Markov chains of interest.*

In this paper we illustrate a new approach to the coupling *technique, which we call* path coupling, *for bounding mixing rates. Previous applications of coupling have required detailed insights into the combinatorics of the problem at hand, and this complexity can make the technique extremely difficult to apply successfully. Path coupling helps to minimize the combinatorial difficulty and in all cases provides simpler convergence proofs than does the standard coupling method. However, the true power of the method is that the simplification obtained may allow coupling proofs which were previously unknown, or provide significantly better bounds than those obtained using the standard method.*

We apply the path coupling method to several hard combinatorial problems, obtaining new or improved results. We examine combinatorial problems such as graph colouring and TWICE-SAT, *and problems from statistical physics, such as the antiferromagnetic Potts model and the hard-core lattice gas model. In each case we provide either a proof of rapid mixing where none was known previously, or substantial simplification of existing proofs with consequent gains in the performance of the resulting algorithms.*

1 Introduction

Exact combinatorial counting is known to be extremely difficult [17], and therefore establishing the approximation complexity of problems in this domain is a significant challenge to theoretical computer science. Even deterministic approximation seems problematic for most problems. However, several #P-hard combinatorial counting problems have been shown to possess a fully polynomial *randomized* approximation scheme (*fpras*). See, for example, [2, 9, 13]. Many of these algorithms are based on establishing polynomial-time convergence of a Markov chain defined on the set of combinatorial objects under consideration. This property is known as *rapid mixing*. Unfortunately, there are few general techniques for proving that a particular chain has this attribute. (See [12] for a recent survey.) A method which has lately proved successful in some situations is the method of *coupling*, a long established technique in applied probability. See, for example, [6, 8, 11]. However, the application of this approach (and it competitors) in this setting has necessarily been somewhat *ad hoc*, requiring detailed combinatorial insights into the target problem.

In this paper we describe a general approach to applying the coupling method to certain Markov chains, which we call *path coupling*. The power of the path coupling method is that it requires only comparisons between adjacent states, rather than arbitrary states, and this results in much simpler analyses. While we stress that the methodology is rather more general than our main theorems here, in Section 2 we apply it to prove two theorems on the convergence rates of Markov chains on product spaces. The first theorem provides the sharpest results, but has the drawback of requiring the Markov chain to be defined on inaccessible as well as accessible states. Fortunately, this is less of a difficulty in some applications than one might expect. The second theorem avoids this difficulty, and allows an easier description of the coupling, but at the expense of some tightness in the bounds obtained. Nevertheless, the results are as tight as the first theorem in some important cases.

In Section 3, we give several non-trivial applications of our theorems to problems from combinatorics and statistical physics. In addition to establishing some new results, we also recover with ease some results that had previously required considerably more ingenuity.

In Section 3.1.1 we give a simple proof of a result of Jerrum [11] on colouring of low-degree graphs. In Section 3.1.2 we give a new and far-reaching extension of this result to hypergraph colourings which, in particular, includes and extends some recent work of Salas and

0272-5428/97 $10.00

Sokal [15] in statistical physics. In Section 3.2 we give a simple proof of a previous result of the authors on SAT instances with two occurrences of each variable. In Section 3.3 we easily extend this to provide a new result on the closely related NOT-ALL-EQUAL-SAT problem. (We will note here that the analysis of this problem had eluded us by the standard coupling approach.) In Section 3.4 we give a new result on counting independent sets of graphs which is related to, but different from, a recent result of Luby and Vigoda [14]. (This result was obtained independently of, and at about the same time as [14].) Finally, in Section 4 we indicate some further applications of the method which will appear elsewhere.

1.1 Notation and preliminaries

Let V and C be finite sets, and define $n = |V|$ and $k = |C|$. For our theorems, we typically consider a finite Markov chain $\mathcal{M}$, with state space $\Omega \subseteq C^V$, the set of functions from V to C, and unique equilibrium distribution π. The reader may find it helpful to keep in mind the example of *proper graph colourings* (we analyse this example in depth in Section 3.1.1): then V is the set of vertices of a graph, and C is a set of colours; the Markov chain $\mathcal{M}$ then has state space the set of functions from vertices to colours, and equilibrium distribution the uniform distribution on the set of *proper colourings*.

In our first theorem, we require that $\Omega = C^V$, but in our second theorem we do not make this assumption. We denote by $D(\mathcal{M})$ the *diameter* of $\mathcal{M}$, i.e. the maximum over all pairs of positive-recurrent states $X, Y \in \Omega$ of the minimum number of transitions necessary to go from X to Y.

For $X \in \Omega$, $v \in V$, and $c \in C$ let us use the notation $X_{v \to c}$ to denote the state resulting from making the transition at X associated with the pair (v, c). Thus

$$X_{v \to c}(w) = \begin{cases} c & \text{if } w = v, \text{ and} \\ X(w) & \text{otherwise.} \end{cases}$$

Using this notation, we may more precisely define the transition structure of $\mathcal{M}$. We first pick $v \in V$ from a fixed distribution J on V. Then we pick $c \in C$ according to a distribution $\kappa_{X,v}$ on C, dependent only on the current state X and v, and make the transition to $X_{v \to c}$. We assume that $X_{v \to c} \notin \Omega$ implies that $\kappa_{X,v}(c) = 0$.

The technique which we use to prove our theorems is known as *coupling*. In particular, we use the "Coupling Lemma" (see e.g. Aldous [1]).

Lemma 1 (Coupling) *Suppose (X, Y) is a random process (the* coupling*) such that marginally, X and Y are both copies of $\mathcal{M}$. Moreover, suppose Y_0 is chosen from π, and μ_t is the distribution of X_t, then*

$$d_{TV}(\mu_t, \pi) \leq \mathbf{P}(X_t \neq Y_t)$$

where d_{TV} is the (total) variation distance metric on measures.

When $X_t = Y_t$, we say that X and Y have *coupled*.

If $G = (V, E)$ is a graph, as in some of our applications, we use $v \sim w$ for the adjacency relation, $N(v) = \{w \in V : v \sim w\}$ for the neighbours of v, and $\delta(v) = |N(v)|$ for the degree of v. We write $\Delta = \max_{v \in V} \delta(v)$ for the maximum degree of G.

2 Two convergence theorems

The essence of the path coupling methodology is very simple. We see, from the Coupling Lemma, that our goal is to construct a joint process on two copies of a Markov chain that will have a probabilistic tendency to come together "quickly". Traditional coupling techniques consider all pairs of states, and show that for most—or all—such pairs, there is a tendency for the two copies of the Markov chain to come closer together (under some metric) in some small (i.e. polynomially bounded) number of steps.

With path coupling, we abstract quickly away from our consideration of *all* pairs of states, by defining a *path*, or sequence of states between an arbitrary pair of states. We then only need to consider pairs of states that are adjacent in some path. Note that states that are adjacent on a path are not necessarily adjacent states in the Markov chain—although they are in the theorems in this section.

If we can show that for all pairs of path-wise adjacent states, that two Markov chains (with an appropriate coupling and metric) started in those two states will come closer together in expectation, then by linearity of expectation and the triangle inequality we may conclude that the entire path is contracting in expectation. A simple induction will thus conclude a proof of rapid mixing.

In the two theorems in this section, we consider a class of Markov chains in which the (not necessarily unique) choice of path arises naturally: we consider Markov chains with state space (some subset of) the set of functions from V to C. Our paths will be constructed simply by insisting that adjacent states on a path differ in their mapping of at most one $v \in V$. The statement of the theorems may seem technical, but the proofs are quite elementary.

The metric that we shall use in the first theorem is Hamming distance, which for states X and Y we shall denote $H(X, Y)$, i.e. $H(X, Y)$ is simply the number of $v \in V$ such that $X(v) \neq Y(v)$. Thus adjacent states on a path have unit Hamming distance.

In both of the theorems, β will be an upper bound on the expected distance between adjacent states after a single time-step.

Theorem 1 (General Path Coupling) *Let* $\Omega = C^V$, *and*

$$\beta = \max_{X,Y\in\Omega, i\in V}\left\{1 - J(i) + \sum_{j\in V} J(j)\mathrm{d_{TV}}(\kappa_{X,j},\kappa_{Y,j})\,\middle|\, Y = X_{i\to c} \text{ for some } c\in C, \text{ and } Y\neq X\right\}.$$

Then, if $\beta < 1$ *and* $t \geq \lceil \ln(n\varepsilon^{-1})/\ln\beta^{-1}\rceil$, *we have* $\mathrm{d_{TV}}(\mu_t,\pi)\leq\varepsilon$.

Proof: Suppose σ_1 and σ_2 are distinct probability distributions on C. Then define the probability distribution $(\sigma_1-\sigma_2)^+$ by

$$(\sigma_1-\sigma_2)^+(c) = \frac{\max\{0,\sigma_1(c)-\sigma_2(c)\}}{\mathrm{d_{TV}}(\sigma_1,\sigma_2)}$$

For $X,Y\in C^V$, let $H(X,Y)$ denote their Hamming distance. Let $h = H(X,Y)$ and let $X = Z_0,Z_1,\ldots,Z_h = Y$ be any sequence in C^V such that $H(Z_{a-1},Z_a) = 1$ (for $a = 1,2,\ldots,h$). Clearly these states are all distinct.

Define the coupling for $\mathcal{M}$ at state (X,Y) by selecting the next state (X',Y') according to the following experiment.

1. Choose $v\in V$ according to J and $c_0\in C$ according to $\kappa_{Z_0,j}$.
2. For $a = 1,2,\ldots,h$ in turn, with probability $\kappa_{Z_a,j}(c_{a-1})/\kappa_{Z_{a-1},j}(c_{a-1})$, let $c_a = c_{a-1}$, otherwise pick c_a according to $(\kappa_{Z_a,j}-\kappa_{Z_{a-1},j})^+$.
3. Make the transition to (X',Y'), where $X' = X_{v\to c_0}$, and $Y' = Y_{v\to c_h}$.

Observe that, marginally, we choose c_a according to $\kappa_{Z_a,j}$, and that this is the "maximal coupling" (i.e. best possible coupling) between $\kappa_{Z_{a-1},j}$ and $\kappa_{Z_a,j}$. In particular $\mathbf{P}(c_a\neq c_{a-1}) = \mathrm{d_{TV}}(\kappa_{Z_a,j},\kappa_{Z_{a-1},j})$.

We shall use Z'_a as an abbreviation of $(Z_a)_{v\to c_a}$, so $X' = Z'_0$, $Y' = Z'_h$. Suppose, in the coupling procedure, that Z_{a-1} and Z_a differ only at i. Then

$$\begin{aligned}
\mathbf{E}\left(H\left(Z'_a,Z'_{a-1}\right)\right) &= 1-\mathbf{P}\left(H\left(Z'_a,Z'_{a-1}\right)=0\right)\\
&\quad+\mathbf{P}\left(H\left(Z'_a,Z'_{a-1}\right)=2\right)\\
&= 1-J(i)\mathbf{P}(c_a = c_{a-1}|v=i)\\
&\quad+\sum_{j\neq i}J(j)\mathbf{P}(c_a\neq c_{a-1}|v=j)\\
&= 1-J(i)(1-\mathrm{d_{TV}}(\kappa_{Z_a,i},\kappa_{Z_{a-1},i}))\\
&\quad+\sum_{j\neq i}J(j)\mathrm{d_{TV}}(\kappa_{Z_a,j},\kappa_{Z_{a-1},j})\\
&\leq \beta.
\end{aligned}$$

Thus $\mathbf{E}(H(X',Y')) \leq \mathbf{E}\left(\sum_{a=1}^h H\left(Z'_a,Z'_{a-1}\right)\right) = \sum_{a=1}^h \mathbf{E}\left(H\left(Z'_a,Z'_{a-1}\right)\right) \leq \sum_{a=1}^h \beta = \beta H(X,Y)$.

Thus, if X_t, and Y_t are the positions of X and Y after t steps, we have that $\mathbf{E}(H(X_t,Y_t)) \leq \beta^t n$. Furthermore, since H is a non-negative integer valued function, $\mathbf{P}(X_t\neq Y_t)\leq\beta^t n$. Applying the Coupling Lemma, we see that $\mathrm{d_{TV}}(\mu_t,\pi)\leq\beta^t n$. Taking logarithms and rearranging establishes the theorem. □

Remark: Suppose we have only that $\beta\leq 1$: then the above theorem appears to tell us nothing about the convergence of the chain. However, the same proof shows that at each step, $H(X,Y)$ cannot increase in expectation, and its value can change either by zero or by one at each step. Suppose that the probability of its value changing at each step is bounded below by α. Then the expected time for the processes to couple, i.e. for $H(X,Y)$ to reach zero, is bounded above by the expected time for a symmetric random walk on the integers $\{0,1,\ldots,n\}$, started at n and with probability α of moving to an adjacent integer, to reach zero. This is $\alpha^{-1}(n^2+n)/2$. Using Markov's inequality, we see that the probability that X and Y have not coupled by time t is bounded above by $\alpha^{-1}(n^2+n)/2(t+1)$. In particular, in order to ensure that the probability that we have not coupled is no greater than e^{-1}, it suffices to simulate $\tau = \lceil e\alpha^{-1}(n^2+n)/2\rceil - 1$ steps of $\mathcal{M}$. Since we may run successive, independent coupling "trials" of length τ, in order to ensure that the probability that we have not coupled is bounded above by ε, it suffices to simulate $\mathcal{M}$ for $\lceil\ln(\varepsilon^{-1})\rceil\tau$ steps. Thus we will have rapid mixing whenever we can show that α^{-1} is polynomial in n and c. We will not give general conditions here for this to be true, but we will consider this observation below. ○

The problem with Theorem 1 is that it requires $\Omega = C^V$. We will relax this assumption by considering a slightly different coupling, for the particularly important class of "Metropolis" Markov chains. The transitions here are as follows. Choose the desired stationary distribution, π. Pick $v\in V$, as before, according to some fixed distribution J. Pick $c\in C$ uniformly at random. Then, with probability $A_{X,v}(c) = \min\{1,\pi(X_{v\to c})/\pi(X)\}$, *accept* and make the transition to $X_{v\to c}$; otherwise *reject* and remain at X. This procedure determines the distributions $\kappa_{X,v}$.

The metric that we use in the following theorem is slightly different from the first. In this theorem, we will use the *minimum transition distance*, which we shall denote $H'(X,Y)$. This is simply the minimum number of transitions of the Markov chain that could be performed in order to move from X to Y.

Theorem 2 (Metropolis Path Coupling) *Let* $\Omega\subseteq C^V$.

For states $Y \neq X$, $Y = X_{i\to c}$ in Ω, let

$$S(X,Y) = \frac{1}{k}\sum_{c\in C} \min\{A_{X,i}(c), A_{Y,i}(c)\}$$

$$F_j(X,Y) = \frac{1}{k}\sum_{c\in C} |A_{X,j}(c) - A_{Y,j}(c)| \quad (j \neq i).$$

Also define

$$\beta = \max_{X,Y\in\Omega, i\in V}\left\{1 - J(i)S(X,Y) + \sum_{j\neq i} J(j)F_j(X,Y)\right| \\ Y = X_{i\to c} \text{ for some } c \in C \text{ such that } Y \neq X\Big\},$$

$$\eta = \min\Big\{\min_{X\in\Omega, i\in V, c\in C}\{A_{X,i}(c) \mid A_{X,i}(c) > 0\}, \\ \min_{X,Y\in\Omega, i\in V, c\in C}\{A_{X,i}(c) - A_{Y,i}(c) \mid A_{X,i}(c) > A_{Y,i}(c)\}\Big\}.$$

1. *If* $\beta < 1$, *then* $d_{TV}(\mu_t, \pi) \leq \varepsilon$ *provided* $t \geq \lceil \ln(D\varepsilon^{-1})/\ln\beta^{-1}\rceil$;
2. *If* $\beta \leq 1$, *then* $d_{TV}(\mu_t, \pi) \leq \varepsilon$ *provided* $t \geq \lceil \ln(\varepsilon^{-1})\rceil\lceil e\eta^{-1}k(D^2+D)/\min_{i\in V}\{J(i)\} - 1\rceil$.

Proof: This proof is very similar to that of the General Path Coupling Theorem, although we use a different coupling here.

For two states, $X, Y \in \Omega$, let $H'(X,Y)$ be the minimum number of transitions required to move from X to Y. Observe that H' is a metric. We will let $h' = H'(X,Y)$. Let $X = Z_0, Z_1, \ldots, Z_{h'}$ be such a minimal sequence of transitions, and note that all these states are distinct.

Suppose we have $(h'+1)$ instances of $\mathcal{M}$, with current states $Z_0, Z_1, \ldots,$ and, $Z_{h'}$, which will evolve jointly by the following experiment:

1. Choose $v \in V$ according to J and $c \in C$ uniformly at random.
2. Choose W uniformly from $[0,1]$. For $a = 0, 1, \ldots, h'$, if $A_{Z_a,v}(c) \geq W$, (accept and) move from Z_a to $(Z_a)_{v\to c}$, otherwise (reject and) remain at Z_a.

This defines a coupling on X and Y. We use Z'_a to denote the state moved to from Z_a in the above experiment.

Observe, by the assumption on transitions of $\mathcal{M}$, that Z_{a-1} and Z_a differ for exactly one element of V, say i. Then $\mathbf{E}\left(H'\left(Z'_{a-1}, Z'_a\right)\right) = 1 + \mathbf{P}\left(H'\left(Z'_{a-1}, Z'_a\right) = 2\right) - \mathbf{P}\left(H'\left(Z'_{a-1}, Z'_a\right) = 0\right)$, since the coupling procedure ensures that $H'\left(Z'_{a-1}, Z'_a\right)$ takes only values in $\{0,1,2\}$.

Now, $H'\left(Z'_{a-1}, Z'_a\right) = 0$ only if we choose $v = i$ in the coupling procedure, and both Z_{a-1} and Z_a accept. Thus we have that $\mathbf{P}\left(H'\left(Z'_{a-1}, Z'_a\right) = 0\right) = J(i)S\left(Z_{a-1}, Z_a\right)$.

The event $H'\left(Z'_{a-1}, Z'_a\right) = 2$ can occur only if we choose $v \in V$ with $v \neq i$, and we accept for precisely one of Z_{a-1} and Z_a. Thus $\mathbf{P}\left(H'\left(Z'_{a-1}, Z'_a\right) = 2\right) = \sum_{j\neq i} J(j)F\left(Z_{a-1}, Z_a\right)$. Thus we have

$$\begin{aligned}\mathbf{E}\left(H'\left(X', Y'\right)\right) &\leq \mathbf{E}\left(\sum_{a=1}^{h'} H'\left(Z'_{a-1}, Z'_a\right)\right) \\ &\quad \text{(since } H' \text{ is a metric)} \\ &\leq \sum_{a=1}^{h'} \mathbf{E}\left(H'\left(Z'_{a-1}, Z'_a\right)\right) \\ &\quad \text{(by linearity of expectation)} \\ &\leq \beta H'(X,Y) \quad \text{(by definition of } \beta).\end{aligned}$$

We have, therefore, that $\mathbf{E}(H'(X_t, Y_t)) \leq \beta^t D$, and, since H' is a non-negative integer valued function, we thus have $\mathbf{P}(X_t \neq Y_t) \leq \beta^t D$. Applying the Coupling Lemma, we see that the variation distance from equilibrium after t steps is bounded above by $\beta^t D$. Taking logarithms and rearranging establishes the first part of the theorem.

To establish the second part of the theorem, we will assume that all acceptance probabilities $A_{X,v}(c) \leq 1/2$ in $\mathcal{M}$. If this is not the case, we may simply halve all the acceptance probabilities to make it so. This is equivalent to having a "do nothing with probability $1/2$" condition at the beginning of each step, and at most doubles the expected number of steps for $\mathcal{M}$ to couple. We allow for this in our calculations. Note that, if $\beta \leq 1$ originally, this will still hold in the revised chain.

Consider the sequence of values taken by $H'(X,Y)$ under the coupling. Assuming $\beta \leq 1$, then at each step, $H'(X,Y)$ cannot increase in expectation, and its value may change either by zero or by one Suppose that the probability that its actual value changes is bounded below by a. Then the expected time for the processes to couple is bounded above by the expected time for a symmetric random walk on the integers $\{0, 1, \ldots, D\}$, with probability a of moving to an adjacent state, started at D, to reach zero. This is $a^{-1}(D^2+D)/2$.

Using Markov's inequality, we see that the probability that we have not coupled by time t is bounded above by $a^{-1}(D^2+D)/2(t+1)$. In particular, in order to ensure that the probability that we have not coupled is no greater than e^{-1}, it suffices to simulate $\tau = \lceil ea^{-1}(D^2+D)/2\rceil - 1$ steps of $\mathcal{M}$. Since we may run successive, independent coupling "trials" of length τ, in order to ensure that the probability that we have not coupled is bounded above by ε, it suffices to simulate $\mathcal{M}$ for $\lceil \ln(\varepsilon^{-1})\rceil\tau$ steps.

It remains to show, however, that $a \geq \eta \min_{i\in V}\{J(i)\}/k$. We do not do this for the coupling above, but for one which is subtly different. Suppose there

is a pair of states, X and Y, such that any choice of $v \in V$, $c \in C$, $W \in [0,1]$ in the previous coupling would result in $H'(X,Y)$ remaining unchanged—we will call this a "stuck pair". If there is no such pair then $H'(X',Y') \neq H'(X,Y)$ with probability at least $\eta \min_{i \in V}\{J(i)\}/k$, and we are done.

Let i be the element of V that is different in X and Z_1, and let $c_1 = Z_1(i)$. Then $X(i) \neq c_1$, since Z_1 is part of a minimal transition sequence from X to Y. Also $c_1 \neq Y(i)$, otherwise choosing i and $Y(i)$ would reduce $H'(X,Y)$ with probability at least η. Let $A, A' \leq 1/2$ be the acceptance probabilities for $Z_1(i)$ in X and Y, given $v = i$. Clearly $A > 0$. If $A > A'$ then we can have $X' = Z_1, Y' = Y$ with probability at least η and we are done. Also if $A < A'$, we can have $X' = X, Y' = Y_{v \to c_1} = Y^*$, say, with probability at least η, and hence we must have $H'(X,Y^*) = H'(X,Y)$. If $A = A'$, then we only have $H'(X,Y^*) \leq H'(X,Y) + 1$.

Our coupling will then be modified as follows. If X and Y are not stuck or $v \neq i$ or $c \neq c_1$, use the previous coupling. Otherwise, accept in X if $W \leq A$ and in Y if $(1 - W) \leq A'$. Note that X and Y are still faithful copies of $\mathcal{M}$. Now, conditional on i and c_1, we have

$$(X',Y') = \begin{cases} (Z_1, Y) & \text{with probability } A, \\ (X,Y) & \text{with probability } (1 - A - A'), \\ (X,Y^*) & \text{with probability } A'. \end{cases}$$

Thus $\mathbf{P}(H'(X',Y') \neq H'(X,Y)) \geq A \geq \eta$. If $A < A'$, then $\mathbf{E}(H'(X',Y')) - H'(X,Y) = -A < 0$. If $A = A'$, then $\mathbf{E}(H'(X',Y')) - H'(X,Y) \leq -A + A' = 0$. So, in either case, we preserve the condition that H' does not increase in expectation. But the probability that it changes is now at least $\eta \min_{i \in V}\{J(i)\}/k$, as required.

Applying the Coupling Lemma completes the proof. □

It should be noted the definitions of β in the two path coupling theorems may yield different values when maximized over adjacent states $X, Y \in \Omega$. In general the first is a lower bound on the second, and this inequality can be strict. They do however coincide in the case $k = 2$, as may be verified by easy calculations.

3 Applications

Here we examine several applications. We consider only the rapid mixing of the relevant Markov chain. Details of the associated approximate counting schemes are omitted, but see, e.g., [6, 11, 12] for the necessary ideas. We note that for all of the examples of chains for which we demonstrate rapid mixing, the associated (exact) counting problem is #P-hard.

3.1 Graph colourings and the Potts model

3.1.1 Graph colourings. For a graph $G = (V,E)$ and set (of colours) C, a function $\chi : V \to C$ is said to be a *colouring*. We will say a vertex v is properly coloured if $\chi(v) \neq \{\chi(w) : w \sim v\}$. A *proper colouring* of G is a colouring with all vertices properly coloured.

Jerrum [11] exhibits a fully polynomial almost uniform sampler for k-colourings of a graph, provided that $k \geq 2\Delta + 1$.

Consider the Markov chain $\mathcal{M}$ with state space the set of all colourings of G and transitions, at state X, defined as follows.

1. Choose v uniformly at random from V, and c uniformly at random from C.

2. If v is properly coloured in $X_{v \to c}$, then $X' = X_{v \to c}$, otherwise $X' = X$.

This is an extension of Jerrum's chain to all of C^V. It is easy to show that the positive-recurrent states of $\mathcal{M}$ are the proper colourings of G, and that the chain is ergodic on these states, but we omit these details here. Let us apply Theorem 1. In this instance, J is the constant function with value $1/n$, and unless $i \sim j$ or $i = j$, $\kappa_{X,j} = \kappa_{Y,j}$. In the case $i = j$, we have $d_{TV}(\kappa_{X,j}, \kappa_{Y,j}) \leq \Delta/k$, since any colour choice that would be accepted in X would also be accepted in Y. Furthermore, for $j \sim i$, $d_{TV}(\kappa_{X,j}, \kappa_{Y,j}) \leq 1/k$, since every colour that would be accepted in X (resp. Y), except possibly $Y(i)$ (resp. $X(i)$), would be accepted in Y (resp. X). Thus $\beta \leq 1 - (1 - \Delta/k)/n + \sum_{j \sim i} 1/kn \leq 1 - 1/n + 2\Delta/kn$.

Thus, applying Theorem 1, we see that $\mathcal{M}$ is rapidly mixing for $k \geq 2\Delta + 1$, and thus comes within ε of its stationary distribution after at most $\lceil \ln(n\varepsilon^{-1})/\ln((kn - k + 2\Delta)/kn) \rceil$ steps.

However, we can possibly do better than this. We need not take J to be constant, and in general, $\beta \leq 1 - J(i)(1 - \delta(i)/k) + \sum_{j \sim i} J(j)/k$. So, if for example we take $J(i) = \delta(i)/2m$, where $m = |E|$, we see that chain would be rapidly mixing provided $k > \max_{v \in V}\{\delta(v) + \sum_{w \sim v} \delta(w)/\delta(v)\}$, i.e. the largest degree of a vertex plus the average degree of its neighbours. This will be less than 2Δ unless there is a vertex of maximum degree for which all of its neighbours are also of maximum degree.

It is natural to inquire whether we can choose J optimally for a particular graph, and it is not hard to see that we may in fact do so. Answering the query, "is there a J such that $\mathcal{M}$ is rapidly mixing for k colours?" is equivalent to checking that a particular linear program is feasible, and finding such a J is equivalent to finding a feasible solution.

It is also possible to show convergence in the case $k = 2\Delta$, as noted in [11] using the remark following Theorem 1. The only stuck pairs are certain proper colourings X, Y with $H(X,Y) = n$. If we simply modify the coupling

to allow them to evolve independently, they cannot remain stuck for long. We omit the details here, and simply note that in this case the mixing rate is $O(kn^3 \log \varepsilon^{-1})$.

3.1.2 Hypergraph colouring and the extended Potts framework. The following subsumes and generalizes section 3.1.1, but with a slightly more complicated approach.

Recall that a hypergraph is a set of vertices, together with a set of "edges"—each edge can contain any number of vertices, but to avoid trivialities, we shall assume here that they each contain at least two. (Clearly, if they all contain exactly two we just have a graph.)

Suppose $G = (V,E)$ is a hypergraph. We will write $v \sim w$ if $\{v,w\} \in e$, for some $e \in E$, and $N(v) = \{w : v \sim w\}$. We consider four degree functions. A vertex $v \in V$ will have degree

$$\begin{aligned}
\delta_1(v) &= \max_{S \subseteq E}\{|S| \mid \forall a,b \in S, a \neq b \Rightarrow a \cap b = \{v\}\},\\
\delta_2(v) &= |\{e \in E | v \in e\}|,\\
\delta_3(v) &= |N(v)|,\\
\delta_4(v) &= \max_{S \subseteq N(V)} \{|S| \mid \forall a,b \in S, e \in E, \{a,b,v\} \not\subseteq e\}.
\end{aligned}$$

Then G is said to have degree $\Delta_i = \max_{v \in V} \delta_i(v)$ $(i = 1,2,3,4)$. All of these definitions of degree, and the notion of k-colourability are in accord with those for graphs. Note that for every vertex, v, $\delta_1(v) \leq \delta_4(v) \leq \min\{\delta_2(v), \delta_3(v)\}$, and hence $\Delta_1 \leq \Delta_4 \leq \min\{\Delta_2, \Delta_3\}$.

The definition of degree δ_1 is standard, and may be found, for example, in Berge [3] or Tomescu [16].

We also define the *co-degree* of two distinct vertices, by $\delta(i,j) = |\{e \in E \mid \{i,j\} \subseteq e\}|$.

As for graphs, we define a colouring to be a function $\chi : V \to C$. The number of *flaws* $f(\chi)$ of χ is the number of edges, $e \in E$, for which all of its vertices are the same colour. A *proper colouring* is a colouring with no flaws. We say colour c is *critical* for $v \in V$ if there is any edge where v is the only vertex not coloured c. We call such an edge a *critical edge*. A hypergraph is k-colourable if it can be properly coloured with k colours.

Many systems in statistical physics are referred to as *anti-ferromagnetic systems*, where less energy is required for a system in which adjacent particles have the same (or in some cases similar) states.

The *k-state Potts model* for anti-ferromagnetism assigns one of k "spins" (colours) to each of the vertices in a graph. The *Hamiltonian* $\mathcal{H}$ defines the *energy* of the colouring, and is equal to the number of flaws in the colouring. It should be clear that we may extend this definition to hypergraphs. In this case we will refer to the model as the *extended Potts framework*. It should be noted that the extended Potts framework is not a physical model *per se*, but rather a *framework* within which physical models, such as the Potts model, may be set. Another model which lies within the framework might define a flaw to be a particle which has the same spin as *all* its neighbours.

According to the axioms of statistical mechanics, if an anti-ferromagnetic system is in equilibrium with surroundings at a temperature T, then the probability of observing any particular colouring χ is proportional to $\exp(-\gamma \mathcal{H}(\chi))$, where $\gamma = 1/kT$,[1] and k is a constant. The probability distribution governing the observed colourings is known as the *Gibbs distribution*.

Consider the Metropolis Markov chain $\mathcal{M}$, defined as follows. At colouring X choose a vertex v and colour c uniformly at random. Accept $X_{v \to c}$ with probability $\min\{1, \exp(f(X) - f(X_{v \to c}))\}$. It is easily checked that the stationary distribution is the Gibbs distribution for the extended Potts framework.

With a chain similar to $\mathcal{M}$, Salas and Sokal [15] use Dobrushin's Uniqueness Criterion (see also [4]) to show absence of phase transition for the Potts model (which is essentially the same as proving rapid mixing) provided $k > 2\Delta$. Jerrum [11] observed, without proof, that it would be sufficient to take $k > 2(1 - e^{-\gamma})\Delta$.

We apply Theorem 1 with J uniform. Consider first the case where $i = j$. Assume without loss of generality that $\pi(X) \geq \pi(Y)$, and define $\rho = d_{TV}(\kappa_{X,j}, \kappa_{Y,j})$, to be the probability that X rejects. There can be at most $\delta_1(i)$ colours that have a non-zero probability of rejecting, since at most $\delta_1(i)$ colours can be critical for i. Enumerate these by $c_1, c_2, \ldots, c_r$. Then ρ is bounded above by $\sum_{l=1}^{r}(1 - e^{-\gamma\sigma_l})/k$, where σ_l is the number of edges critical for i with colour c_l. Now, since $1 - e^{-x}$ is a convex increasing function, and $\sum_{l=1}^{r} \sigma_l \leq \delta_2(j)$, we have $\rho \leq \delta_1(i)[1 - e^{-\gamma\delta_2(i)/\delta_1(i)}]/k$.

For the cases in which $j \neq i$, we have $d_{TV}(\kappa_{X,j}, \kappa_{Y,j}) \leq (1 - e^{-\gamma\delta(i,j)})/k$, since there could be most $\delta(i,j)$ additional flaws for any colour choice at j caused by changing only the colour of i. Also there are at most $\delta_4(i)$ vertices for which $d_{TV}(\kappa_{X,j}, \kappa_{Y,j}) > 0$. Any two such vertices must be adjacent to i, but not both in any edge containing i. This follows from the fact that an edge may be critical for at most one vertex other than i in either X or Y.

Thus $\mathcal{M}$ is rapidly mixing provided that

$$\max_{i,j \in V}\left\{1 - \frac{1}{n} + \delta_1(i)\frac{1 - e^{-\gamma\delta_2(i)/\delta_1(i)}}{nk} + \delta_4(i)\frac{1 - e^{-\gamma\delta(i,j)}}{nk}\right\} < 1.$$

[1] The usual symbol is β, but we use γ here to avoid confusion

Thus, in particular, $k > \Delta_1 + \Delta_4$ is a sufficient condition. Also it is possible to show that the only stuck pairs are proper colourings of a *regular graph* of degree Δ, and hence we could show convergence also in the case $k \geq \Delta_1 + \Delta_4$.

In the case of the Potts model, we have $\delta_r = \delta$ $(r = 1,2,3,4)$, and $\delta(i,j) \leq 1$. Thus a sufficient condition is $k > 2\Delta(1 - e^{-\gamma})$. Jerrum [11] mentions that a similar result, for a slightly different Markov chain has been obtained by Salas and Sokal. Note also that once again the equality case can be handled using the remark following Theorem 1.

3.2 Sink-free graph orientations and TWICE-SAT

In TWICE-SAT [6], we consider the set $\Omega \subseteq \{\text{True}, \text{False}\}^V$ of satisfying assignments to a Boolean formula in conjunctive normal form, where V is the set of variables, and each variable appears at most twice in total. This restriction of the familiar SAT problem remains #P-complete in its counting variant.

An *orientation* of an undirected graph, G, is an assignment of a direction to each of the edges of G. A *sink-free orientation* is an orientation in which no vertex has zero out-degree. The problems associated with sink-free graph orientations may be parsimoniously reduced to TWICE-SAT instances [6], so henceforth we consider only the TWICE-SAT problem.

We consider a Markov chain $\mathcal{M}$, on Ω. Transitions on $\mathcal{M}$ are made by choosing $v \in V$ and truth value c uniformly at random. If assigning c to v results in a satisfying assignment, we accept and make a transition to this state. This chain was first shown to be rapidly mixing by the authors [6]. Let m be the number of clauses. Then it is possible to show that, under certain (non-restrictive) conditions [6] on the instance, the chain is ergodic with diameter $\max\{2m, n\}$, but we omit these details here.

We may easily apply the Theorem 2 to this chain. If X and Y are two assignments that differ only at i, write $Y = X_{\bar{i}}$. Consider the formula for β. If $Y = X_{\bar{i}}$ for some X and i in the maximum, then

$$\beta = 1 - 1/n + |\{j \in V \mid j \neq i; \text{ precisely one of } X_{\bar{j}}, Y_{\bar{j}} \in \Omega\}|/2n.$$

Now if only one of $X_{\bar{j}}$ and $Y_{\bar{j}}$ is in Ω, this must be caused by a clause in which both i and j appear as variables. Furthermore, in this clause, the literal of variable j must appear as True, and the literal of i must appear as False in one of X and Y and True in the other. All other literals in the clause must appear as False. Clearly there can be at most one such j in each clause in which i appears, i.e. there are at most two such js. Thus $\beta \leq 1$.

Applying conclusion 2 of the Metropolis Path Coupling Theorem, with $\eta = 1$, we see immediately that $\mathcal{M}$ is rapidly mixing with mixing rate $O(n^3 \log \varepsilon^{-1})$; the best bound known for this Markov chain previously was $O((n^3 + nm^3)\log \varepsilon^{-1})$ [6].

3.3 Sink- and source-free graph orientations

An *orientation* of an undirected graph, G, is an assignment of a direction to each of the edges of G.

A *sink- and source-free orientation* is an orientation of an undirected graph in which every vertex has both positive in-degree and positive out-degree. In the SAT setting of Section 3.2 this corresponds to a restriction of the well-known problem NOT-ALL-EQUAL-SAT. The proof offered in this section is the only known proof of rapid mixing for a Markov chain on this state space.

For a graph $G = (V, E)$, fix an orientation, and call this O. Let $\mathcal{M}$ be a Markov chain on the set of orientations of G. We may regard these orientations as the set of functions from E to $\{A, D\}$, where an edge takes value A in an orientation if it agrees with O, and D if it disagrees. If X is a sink- and source-free orientation of G, then $\mathcal{M}$ will have transitions modelled by picking an edge and one of A and D uniformly at random. If the orientation formed by orienting this edge to the choice of A or D results in a sink- and source-free orientation then accept this, otherwise reject, and remain at the current orientation. It can be shown that, if G has minimum degree at least 3, this chain is ergodic, with diameter at most $m + n$, where $m = |E|$. Again we omit the details.

We will apply the Metropolis Path Coupling theorem to show that $\mathcal{M}$ is rapidly mixing on any graph with minimum degree at least four. The associated counting problem, even for this restricted case, is #P-complete.

If X and Y are two orientations that differ only at i, we shall write $Y = X_{\bar{i}}$. Then

$$\beta = 1 - 1/n + |\{j \in V \mid j \neq i; \text{ precisely one of } X_{\bar{j}}, Y_{\bar{j}} \in \Omega\}|/2n.$$

Now if only one of $X_{\bar{j}}$ and $Y_{\bar{j}}$ is in Ω there must be a vertex, v, incident on both i and j. Furthermore, if j is oriented away from v then all other edges incident on v must be oriented towards it in one of X or Y. Similarly, if j is oriented toward v then all other edges incident on v must be oriented away from it in one of X or Y. If $\delta(v) > 3$ then there can be at most one such j incident on v. (As an aside note that, if $\delta(v) < 3$, then X and Y cannot both be in Ω). Since there are only two vertices incident on i, it follows that provided every vertex in G has degree at least four, $\beta \leq 1$. In this case, Theorem 2, with $\eta = 1$, establishes that $\mathcal{M}$ is rapidly mixing with mixing time $O(m^3 \log \varepsilon^{-1})$.

3.4 Independent sets and conserved hard-core lattice gases

Consider a graph G on n vertices, where each vertex may be either *occupied* (by a single particle) or *unoccupied*. An instance of a conserved hard-core lattice gas on G with s particles is a configuration in which every vertex adjacent to an occupied vertex is unoccupied. The set of occupied vertices in a hard-core lattice gas is (equivalently) an independent set. In this section we describe a fully polynomial almost uniform sampler for instances of a conserved hard-core lattice gas, provided some condition (to be determined) holds on s. Since this problem is equivalent to finding independent sets almost uniformly at random from G^r, we shall restrict our attention to this problem. This is the first proof of rapid mixing for a Markov chain on this state space. A related problem has been considered by Luby and Vigoda [14] (see also Section 4.)

Consider the set of functions from particles $P = \{1,2,\ldots,s\}$ to vertices, V. We call such a function ξ a *placement*. We shall define a Markov chain $\mathcal{M}$ on all placements, to which we shall apply the General Path Coupling Theorem. A particle i is said to be *safe* in ξ if there is no $j \in P$ such that either $\xi(i) = \xi(j)$ or $\xi(i) \sim \xi(j)$.

Transitions of $\mathcal{M}$ will be as follows. Assume the current state is X. Pick $p \in P$ and $v \in V$ uniformly at random. If p would be safe at v, then accept, and move to this new placement, otherwise reject and remain at X.

Then $d_{TV}(\kappa_{X,i}, \kappa_{Y,i}) \leq \sum_{j \neq i}(\delta(X(j)) + 1)/n \leq (s-1)(\Delta+1)/n$, since we only choose a different vertex if one of X or Y rejects, and this can happen only if we choose a vertex occupied by, or adjacent to, another particle.

Now $d_{TV}(\kappa_{X,j}, \kappa_{Y,j}) \leq (\delta(X(i)) + 1 + \delta(Y(i)) + 1)/2n \leq (\Delta+1)/n$. Thus

$$\beta \leq 1 - 1/s + 2(\Delta+1)(s-1)/sn,$$

and a sufficient condition for $\beta < 1$ is $s < n/2(\Delta+1) + 1$.

Applying the General Path Coupling Theorem establishes that we have rapid mixing to the uniform stationary distribution on independent sets of size s, provided that $s < n/2(\Delta+1) + 1$. Counting the number of independent sets of size s remains a #P-complete problem when restricted to $s < n/2(\Delta+1) + 1$.

4 Conclusions and further work

As we have seen, the path coupling method may readily be applied to a variety of cases that are quite different. This has enabled us to prove several new results on sampling combinatorial structures, and where the results have been known previously, we have proofs that are far simpler than their original counterparts.

The path coupling method does not just extend to the examples presented here: the authors [5] have used this technique elsewhere to substantially improve the best bound on sampling almost uniformly from the set of linear extensions of a partial order from $O(n^5 \log n + n^4 \log \varepsilon^{-1})$ to $O(n^3 \log n + n^3 \log \varepsilon^{-1})$. We observe that Ωn^3 is a *lower* bound on the best chains known for this problem.

Luby and Vigoda [14] have recently proved the rapid mixing of a Markov chain for the (unconserved) hard-core model—this differs from the conserved model considered in this paper in that in the unconserved model the number of particles is not fixed. A Metropolis Markov chain with the same stationary distribution as the Luby-Vigoda chain succumbs easily to our general theorems here. The result obtained is incomparable to that obtained by Luby and Vigoda. However, a more careful analysis of the Luby-Vigoda chain by path coupling gives a result which generalizes both of these (see Dyer and Greenhill [10]). The same paper also describes a significant improvement of Luby and Vigoda's result using a modified chain.

Path coupling has also been used by the authors and Greenhill [7] to beat the $k \geq 2\Delta$ bound for k-colouring a graph, for some special cases. In particular, rapid mixing has been established for a Markov chain on the set of 5-colourings for graphs of maximum degree 3, and on the set of 7-colourings for triangle-free 4-regular graphs.

Acknowledgments

We are grateful to Mark Jerrum for many useful interactions. Also, we thank Alan Sokal and Neal Madras for an instructive e-mail tutorial on approaching the Potts model via Dobrushin's criterion. A discussion of this appeared in an earlier version of this paper [4].

References

[1] David Aldous. Random walks on finite groups and rapidly mixing Markov chains. In A. Dold and B. Eckmann, editors, *Séminaire de Probabilités XVII 1981/1982*, volume 986 of *Springer-Verlag Lecture Notes in Mathematics*, pages 243–297. Springer-Verlag, New York, 1983.

[2] Noga Alon, Alan Frieze, and Dominic Welsh. Polynomial time randomized approximation schemes for Tutte-Gröthendieck invariants: The dense case. *Random Structures and Algorithms*, 6(4):459–478, 1995.

[3] Claude Berge. *Graphs and Hypergraphs*, volume 6 of *North-Holland Mathematical Library*, chapter 19, pages 428–447. North-Holland Publishing Company, Amsterdam-London, 1973.

[4] R. Bubley and M. E. Dyer. Path coupling, Dobrushin uniqueness, and approximate counting. Technical Report 97.04, School of Computer Studies, University of Leeds, January 1997.

[5] Russ Bubley and Martin Dyer. Faster random generation of linear extensions, April 1997. (Preprint).

[6] Russ Bubley and Martin Dyer. Graph orientations with no sink and an approximation for a hard case of #SAT. In *Proceedings of the Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 248–257, New Orleans, Louisiana, 5–7 January 1997.

[7] Russ Bubley, Martin Dyer, and Catherine Greenhill. Beating the 2Δ bound for approximately counting colourings: A computer-assisted proof of rapid mixing, 1997. (Preprint).

[8] Russ Bubley, Martin Dyer, and Mark Jerrum. A new approach to polynomial-time generation of random points in convex bodies. Technical report, Department of Computer Science, University of Edinburgh, April 1996. Report ECS-LFCS-96-343.

[9] Martin Dyer, Alan Frieze, and Ravi Kannan. A random polynomial-time algorithm for approximating the volume of convex bodies. *Journal of the ACM*, 38(1):1–17, January 1991.

[10] Martin Dyer and Catherine Greenhill. On Markov chains for independent sets, 1997. (Preprint).

[11] Mark Jerrum. A very simple algorithm for estimating the number of k-colorings of a low-degree graph. *Random Structures and Algorithms*, 7:157–165, 1995.

[12] Mark Jerrum and Alistair Sinclair. The Markov chain Monte Carlo method: an approach to approximate counting and integration. In Dorit Hochbaum, editor, *Approximation Algorithms for NP-Hard Problems*, pages 482–520. PWS Publishing, Boston, 1996.

[13] Richard M. Karp and Michael Luby. Monte-Carlo algorithms for enumeration and reliability problems. In *24th Annual Symposium on Foundations of Computer Science*, pages 56–64, Tucson, Arizona, 7–9 November 1983. IEEE.

[14] Michael Luby and Eric Vigoda. Approximately counting up to four. In *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing*, pages 682–687, El Paso, Texas, 4–6 May 1997.

[15] Jesús Salas and Alan D. Sokal. Absence of phase transition for antiferromagnetic Potts models via the Dobrushin uniqueness theorem. *Journal of Statistical Physics*, 86(3–4):551–579, February 1997.

[16] Ioan Tomescu. Sur le problème du coloriage des graphes généralisés. *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences, Série A*, 267(6):250–252, 5 August 1968.

[17] Leslie G. Valiant. The complexity of enumeration and reliability problems. *SIAM Journal on Computing*, 8(3):410–421, August 1979.