*Bernoulli* **23**(1), 2017, 603–625
DOI: 10.3150/15-BEJ758

# Convergence rates for a hierarchical Gibbs sampler

OLIVER JOVANOVSKI[1] and NEAL MADRAS[2]

[1]*Mathematical Institute, Leiden University, P.O. Box 9512, 2300 RA, Leiden, The Netherlands. E-mail: o.jovanovski@math.leidenuniv.nl*
[2]*Department of Mathematics and Statistics, York University, 4700 Keele Street, Toronto, Ontario M3J 1P3, Canada. E-mail: madras@mathstat.yorku.ca*

We establish results for the rate of convergence in total variation of a particular Gibbs sampler to its equilibrium distribution. This sampler is for a Bayesian inference model for a gamma random variable, whose only complexity lies in its multiple levels of hierarchy. Our results apply to a wide range of parameter values when the hierarchical depth is 3 or 4. Our method involves showing a relationship between the total variation of two ordered copies of our chain and the maximum of the ratios of their respective coordinates. We construct auxiliary stochastic processes to show that this ratio converges to 1 at a geometric rate.

*Keywords:* convergence rate; coupling; gamma distribution; hierarchical Gibbs sampler; Markov chain; stochastic monotonicity

## 1. Introduction

A basic purpose of Markov chain Monte Carlo (MCMC) is to generate samples from a given "target" probability distribution by inventing a Markov chain that has the target as its equilibrium, and then sampling from long runs of this chain. A widely used class of MCMC methods known as Gibbs samplers (see below) is especially well suited for Bayesian statistical models. There is a significant amount of theory showing that a Markov chain satisfying some fairly general conditions (see, for example, [14]) will converge to an equilibrium in distribution, as well as in the stronger measure of total variation. Mere knowledge of convergence is often not enough, and it is of both theoretical and practical interest to consider the rate at which convergence proceeds. Bounds on this rate provide a rigorous degree of certainty as to how far the Markov chain is from its equilibrium distribution, and they aid in assessing the efficiency of this sampling procedure.

Unfortunately, the theoretical methods available for proving such bounds cannot handle many realistic statistical models of even moderate complexity. In particular, general methods for assessing convergence rates in hierarchical Gibbs samplers are scarce, and it was this state of affairs that motivated the present work. Our specific goal was to explore the extent that multiple levels of hierarchy would affect the convergence rate analysis. Accordingly, we focused on a Bayesian model that had several hierarchical levels, but was otherwise as simple as possible (indeed, too simple to be of much use in real applications). We hope that our analysis of this toy model will serve to guide future analyses of Gibbs samplers for more complex multi-level hierarchical models.

Our Bayesian model corresponds to the following scenario. We are given a real number $x > 0$ with the information that it was drawn from a $\Gamma(a_1, u_1)$ distribution, that is, the Gamma distribution with probability density function

$$f(z) = \frac{u_1^{a_1}}{\Gamma(a_1)} z^{a_1 - 1} e^{-z u_1} \qquad (z > 0).$$

Here the shape parameter $a_1 > 0$ is fixed, but the inverse scale parameter $u_1$ is itself the product of random sampling from an independent $\Gamma(a_2, u_2)$ distribution. Once again, we assume that $a_2 > 0$ is a given constant, while $u_2$ is sampled in an analogous manner. This process continues until we reach $u_n \sim \Gamma(a_{n+1}, b)$, where now both $a_{n+1} > 0$ and $b > 0$ are given. The joint density $p$ of $(x, u_1, \dots, u_n)$ is known up to proportionality:

$$p(x, u_1, \dots, u_n) \propto x^{a_1 - 1} \left( \prod_{i=1}^{n} u_i^{a_i + a_{i+1} - 1} \right) \exp\left( \sum_{i=1}^{n+1} -u_i u_{i-1} \right), \tag{1.1}$$

where for convenience we set $u_0 := x$ and $u_{n+1} := b$. Therefore, the resulting posterior distribution of $u = (u_1, \dots, u_n)$ (i.e., given $x$ as well as all other parameters) has the density function

$$\bar{\pi}(u_1, \dots, u_n) \propto \left( \prod_{i=1}^{n} u_i^{a_i + a_{i+1} - 1} \right) \exp\left( \sum_{i=1}^{n+1} -u_i u_{i-1} \right). \tag{1.2}$$

This is the function underlying Bayesian inference for the $u_i$'s given the data $x$. We also see from (1.1) that for $1 \le i \le n$, the conditional distribution of $u_i$ given everything else is

$$u_i | x, u_{j \ne i} \sim \Gamma(a_i + a_{i+1}, u_{i-1} + u_{i+1}).$$

This property will be key to implementing our Gibbs sampler.

Bayesian hierarchical models have been a popular statistical representation used to handle a variety of problems (see, for instance, [4,15,17] or [3,6,8,9]). The Gibbs sampler [7] has been a very popular MCMC algorithm for obtaining a sample from a probability distribution that is difficult to sample from directly. In its fundamental form, this algorithm works on a vector $u$ by selecting (systematically, randomly, or otherwise) one of the vector's components $u_i$ and updating this component only, by drawing from the probability distribution of $u_i$ given $(u_{j \ne i})$. General convergence results have been derived for some Gibbs samplers (e.g., [18]), however due to their limitations it is often not possible to infer quantitative bounds directly from these results.

In this paper, we will focus on our model with $n = 4$, with a short section dedicated to results for the case $n = 3$. The case $n = 3$ actually reduces to a one-dimensional Markov chain, which is relatively tractable. The case $n = 4$ (as well as $n = 5$) reduces to a two-dimensional Markov chain, and hence requires new approaches. For cases $n > 4$, we refer the reader to [10], where we derive similar results under more restrictive constraints on the parameters.

The type of result that we shall prove is the following. Fix $n = 4$. Let $\{G^t : t \ge 0\}$ be the Gibbs sampler for (1.1) that updates the odd coordinates and then the even coordinates at each iteration (this Markov chain is defined explicitly in (1.4) below). Fix the initial point $G^0$. Assume $a_i > 1/2$ for every $i$ (or more generally, assume condition (1.7) below). Then

$$d_{\mathrm{TV}}\big(G^t, \bar{\pi}\big) \le A_1 r^{(t-4)/4d}(a_2 + a_3 + a_4 + a_5) + A_2 \beta^{t/2} \qquad \text{for every } t > 0, \tag{1.3}$$

where $d_{\mathrm{TV}}$ is the total variation metric (described in Section 1.1 below), and $r$, $d$, $\beta$, $A_1$ and $A_2$ depend only on $a_1, \dots, a_5$, $b$, and $x$ (and $G^0$ for $A_1$ and $A_2$), with $r < 1$ and $\beta < 1$. The constants $r$, $d$, and $\beta$ have explicit formulas (listed in Appendix), while each $A_i$ involves an integral with respect to $\bar{\pi}$ that we know how to estimate.

## 1.1. Formulation and reduction of the problem

Our aim is to construct a Gibbs sampler on $\mathbb{R}^4_+$ and show that it converges rapidly to the target distribution with density function given by (1.2) with $n = 4$. For $n > 4$, we use a similar approach in [10].

To describe distance from equilibrium, we use the total variation metric $d_{\mathrm{TV}}$, which is defined as follows. For two probability measures $\mu_1$ and $\mu_2$ on the same state space $\Omega$, define $d_{\mathrm{TV}}(\mu_1, \mu_2) := \inf \mathbb{P}(X_1 \ne X_2)$, where the infimum is over all joint distributions $\mathbb{P}$ of $(X_1, X_2)$ such that $X_1 \sim \mu_1$ and $X_2 \sim \mu_2$. If $Y_i$ denotes a random variable with distribution $\mu_i$, then we shall also write $d_{\mathrm{TV}}(Y_1, Y_2)$ and $d_{\mathrm{TV}}(Y_1, \mu_2)$ for $d_{\mathrm{TV}}(\mu_1, \mu_2)$. It is known (e.g., Chapter I of [11]) that the infimum is achieved by some $\mathbb{P}$, and that we can also express $d_{\mathrm{TV}}(\mu_1, \mu_2)$ as the supremum of $|\mu_1(A) - \mu_2(A)|$ over all measurable $A \subset \Omega$.

***Notation.*** We shall write $\vec{u} = (u_1, u_2, u_3, u_4)$ for points in $\mathbb{R}^4$. We shall often refer to points of $\mathbb{R}^2$ consisting of the second and fourth entries of $\vec{u}$. We shall then omit the $\vec{}$ and write $u = (u_2, u_4)$.

We first consider the Markov chain which sequentially updates its coordinates as follows. For $i \in \{1, 2, 3, 4\}$, let

$$P_i(\vec{v}, d\vec{w}) := \bigg(\prod_{j \ne i} \delta_{v_j}(w_j)\bigg) h_i(w_i | \vec{v}) \, dw_i,$$

where $h_i(\cdot|\vec{v})$ is the $\Gamma(a_i + a_{i+1}, v_{i-1} + v_{i+1})$ density function given $\vec{v}$, and where for convenience we have defined $v_0 := x$ and $v_5 := b$. In other words, $P_i$ is the probability kernel that updates (only) the $i$th coordinate according to the conditional density $h_i$. Now define

$$P := P_1 P_3 P_2 P_4, \tag{1.4}$$

the Gibbs sampler Markov chain that updates the odd coordinates and then the even coordinates. We will show that $P$ converges to equilibrium at a geometric rate, and we will give a bound on the rate of convergence.

It will be useful to represent our Markov chain using iterated random functions [5,12] as follows. Let $\{\gamma_i^t : i = 1, 2, 3, 4; t = 1, 2, \dots\}$ be a collection of independent random variables with each $\gamma_i^t$ having the $\Gamma(a_i + a_{i+1}, 1)$ distribution. Then define the sequence of random functions $\bar{F}^t : \mathbb{R}_+^4 \to \mathbb{R}_+^4$ $(t = 1, 2, \dots)$ by

$$\begin{aligned}\bar{F}^t(\vec{u}) &= \big(\bar{F}_1^t(\vec{u}), \bar{F}_2^t(\vec{u}), \bar{F}_3^t(\vec{u}), \bar{F}_4^t(\vec{u})\big) \\ &= \left(\frac{\gamma_1^t}{x+u_2}, \frac{\gamma_2^t}{\gamma_1^t/(x+u_2)+\gamma_3^t/(u_2+u_4)}, \frac{\gamma_3^t}{u_2+u_4}, \frac{\gamma_4^t}{b+\gamma_3^t/(u_2+u_4)}\right).\end{aligned} \tag{1.5}$$

Then for any initial $\vec{u}^0 \in \mathbb{R}_+^4$, the random sequence $\vec{u}^0, \vec{u}^1, \dots$ defined recursively by $\vec{u}^{t+1} = \bar{F}^{t+1}(\vec{u}^t)$ is a Markov chain with transition kernel $P$.

Observe that $\bar{F}^t(\vec{u})$ does not depend on $u_1$ or $u_3$. It follows that if $\{u^t\}$ is a version of the Markov chain (1.4), then the sequence $\{(u_2^t, u_4^t)\}$ is itself a Markov chain in $\mathbb{R}_+^2$. Accordingly, we define the random functions $F^t : \mathbb{R}_+^2 \to \mathbb{R}_+^2$ $(t = 1, 2, \dots)$ by

$$F^t(u_2, u_4) = \big(F_2^t(u_2, u_4), F_4^t(u_2, u_4)\big) = \left(\frac{\gamma_2^t}{\gamma_1^t/(x+u_2)+\gamma_3^t/(u_2+u_4)}, \frac{\gamma_4^t}{b+\gamma_3^t/(u_2+u_4)}\right).$$

Thus $\bar{F}_i^t(\vec{u}) = F_i^t(u_2, u_4)$ for $i = 2, 4$ and all $\vec{u} \in \mathbb{R}_+^4$ and all $t$. Moreover, the Markov chain $\{(u_2^t, u_4^t)\}$ is given by the random recursion

$$\big(u_2^{t+1}, u_4^{t+1}\big) = F^{t+1}\big(u_2^t, u_4^t\big). \tag{1.6}$$

Let $\bar{\pi}$ be the probability measure on $\mathbb{R}_+^4$ with density function (1.2). Then it is well known (see, e.g., Section 2.3 of [1]) that $\bar{\pi}$ is the equilibrium distribution of the Markov chain defined by (1.4). It follows that the marginal distribution of the even coordinates of $\bar{\pi}$, which we denote by $\pi$, is the equilibrium distribution of (1.6). Furthermore, the following simple argument illustrates that it suffices to bound the distance to equilibrium of (1.6).

**Lemma 1.1.** *Let $\vec{\Upsilon}^t$ be a copy of the Markov chain (1.4) on $\mathbb{R}_+^4$ and let $\Phi^t$ be a copy of (1.6) on $\mathbb{R}_+^2$. Assume $(\Upsilon_2^0, \Upsilon_4^0) = \Phi^0$, that is, the initial conditions agree. Then $d_{\mathrm{TV}}(\vec{\Upsilon}^{t+1}, \bar{\pi}) \le d_{\mathrm{TV}}(\Phi^t, \pi)$.*

**Proof.** Fix $t$. There exists a jointly distributed pair of random vectors $(\Psi, \Lambda)$ with $\Psi = (\Psi_2, \Psi_4) \sim \Phi^t$ and $\Lambda = (\Lambda_2, \Lambda_4) \sim \pi$ such that $d_{\mathrm{TV}}(\Phi^t, \pi) = \mathbb{P}[\Psi \ne \Lambda]$. Then $\vec{\Upsilon}^{t+1} \sim \bar{F}^{t+1}(1, \Psi_2, 1, \Psi_4)$ and $\bar{F}^{t+1}(1, \Lambda_2, 1, \Lambda_4) \sim \bar{\pi}$. Hence,

$$\begin{aligned}d_{\mathrm{TV}}\big(\vec{\Upsilon}^{t+1}, \bar{\pi}\big) &\le \mathbb{P}\big[\bar{F}^{t+1}(1, \Psi_2, 1, \Psi_4) \ne \bar{F}^{t+1}(1, \Lambda_2, 1, \Lambda_4)\big] \\ &\le \mathbb{P}[\Psi \ne \Lambda] \\ &= d_{\mathrm{TV}}\big(\Phi^t, \pi\big).\end{aligned}$$ □

We can now state our main results. Let $\mathcal{U}^t$ and $\mathcal{W}^t$ be two copies of the Markov chain (1.6) starting at points $\mathcal{U}^0$ and $\mathcal{W}^0$ respectively. We define the condition

$$a_1 + a_4 > 1, \qquad a_2 + a_5 > 1, \qquad a_2 + a_3 > 1, \qquad a_3 + a_4 > 1, \qquad a_4 + a_5 > 1. \tag{1.7}$$

Let $M = \max\{\mathcal{U}_2^0, \mathcal{U}_4^0, \mathcal{W}_2^0, \mathcal{W}_4^0\}$ and $m = \min\{\mathcal{U}_2^0, \mathcal{U}_4^0, \mathcal{W}_2^0, \mathcal{W}_4^0\}$, and define $R_0 = \frac{M}{m}$ and $J_0 = 2m + 1/(2m)$. The constants $\eta$, $d$, $\beta$ and $r$ appearing in the statement of Theorem 1.1 are defined in Appendix (as well as in Section 4), and depend only on the parameters $x, b, a_1, \dots, a_5$.

**Theorem 1.1.** *Assume that* (1.7) *holds, and fix* $\mathcal{U}^0$ *and* $\mathcal{W}^0$. *If* $J_0 < \eta$, *then for* $t > 0$,

$$d_{\mathrm{TV}}\big(\mathcal{U}^{t+3}, \mathcal{W}^{t+3}\big) \le 3r^{t/2d}(a_2 + a_3 + a_4 + a_5)(R_0 - 1).$$

*For general values of* $J_0$, *we have*

$$d_{\mathrm{TV}}\big(\mathcal{U}^{t+3}, \mathcal{W}^{t+3}\big) \le 3r^{t/4d}(a_2 + a_3 + a_4 + a_5)(R_0 - 1) + \frac{\max\{J_0, \eta\}}{\eta}\beta^{\lfloor t/2\rfloor+3}.$$

We explain our need for condition (1.7) in Section 4. Taking $\mathcal{W}^0 \sim \pi$ leads to the following.

**Corollary 1.1.** *Assume that* (1.7) *holds, and fix* $\mathcal{U}^0$. *For* $t > 0$,

$$d_{\mathrm{TV}}\big(\mathcal{U}^{t+3}, \pi\big) \le 3r^{t/4d}(a_2 + a_3 + a_4 + a_5)\mathbb{E}_\pi[R_0 - 1] + \left(\frac{\mathbb{E}_\pi[J_0]}{\eta} + 1\right)\beta^{\lfloor t/2\rfloor+3}.$$

The quantities $\mathbb{E}_\pi[R_0]$ and $\mathbb{E}_\pi[J_0]$ depend only on $\pi$ and $\mathcal{U}^0$, and can be estimated with a bit of effort. This is done in Appendix B of [10] for the case $\mathcal{U}^0 = (1, 1)$ (see also the end of Section 5 in the present paper). Observe that equation (1.3) follows from Corollary 1.1 and Lemma 1.1.

While the above bounds are of theoretical interest for their explicit nature, their practical value lies more in their qualitative assurance of exponentially rapid convergence rather than in the numerical values of the bounds (as computed in Sections 5 and 6).

## 1.2. Outline of the paper

The proof of Theorem 1.1 relies on two kinds of coupling constructions. In Section 2, we consider a partial order "$\preceq$" on $\mathbb{R}_+^2$ that is preserved by a natural coupling using the recursion (1.6). We use this construction, in which the two chains $\mathcal{U}^t$ and $\mathcal{W}^t$ of Theorem 1.1 stay between two other chains $u^t$ and $w^t$ that are also copies of (1.6), up until a "one-shot coupling" time at which $u^t$ and $w^t$ try to coalesce (succeeding with high probability, desirably). In Section 3, we define the stochastic process $R_t$ which is an upper bound for the ratio $\max_i\{\frac{w_i^t}{u_i^t}\}$, and we show that the rate of convergence of $R_t \to 1$ can be related to the rate at which (1.6) converges to equilibrium. But the rate at which $R_t$ approaches 1 does depend on the size of the values $u_2^t$ and $u_4^t$, and we show that if often enough these two values are neither too large nor too small, then $R_t \to 1$ at a geometric rate. To fulfill this condition, we introduce a number of auxiliary processes in Section 4 that provide upper bounds for the terms $\{u_2^t, u_4^t, \frac{1}{u_2^t}, \frac{1}{u_4^t}\}$, and we show that they are frequently less than a fixed constant. Section 4 culminates in the proof of Theorem 1.1. In Section 5, we prove Corollary 1.1 and discuss the evaluation of the terms $\mathbb{E}_\pi[R_0]$ and $\mathbb{E}_\pi[J_0]$ that appear in the corollary. Section 6 addresses the case $n = 3$.

## 2. Monotone coupling and one-shot coupling

For $p = (p_2, p_4) \in \mathbb{R}_+^2$ and $q = (q_2, q_4) \in \mathbb{R}_+^2$, define the partial order $p \preceq q$ to mean $p_2 \le q_2$ and $p_4 \le q_4$. Given several initial points $p^0, q^0, \dots$ (possibly random) in $\mathbb{R}_+^2$, we can produce several versions of the Markov chain (1.6) using $p^{t+1} = F^{t+1}(p^t)$, $q^{t+1} = F^{t+1}(q^t)$, and so on (crucially, we use the same random variables $\{\gamma_i^t\}$ in each version). We refer to this as the "uniform coupling." This coupling is clearly monotone, in the sense that if $p^0 \preceq q^0$ then $p^t \preceq q^t$ for all times $t$.

Our primary objective is to obtain upper bounds on $d_{\mathrm{TV}}(\mathcal{U}^t, \mathcal{W}^t)$, where $\{\mathcal{U}^t\}$ and $\{\mathcal{W}^t\}$ are two copies of the Markov chain (1.6) with initial points $\mathcal{U}^0$ and $\mathcal{W}^0$ in $\mathbb{R}_+^2$. The initial points could be random, but we will often treat them as fixed either by conditioning on events at $t = 0$ or by explicit assumption. Given the initial points $\mathcal{U}^0$ and $\mathcal{W}^0$, define $m := \min\{\mathcal{U}_2^0, \mathcal{U}_4^0, \mathcal{W}_2^0, \mathcal{W}_4^0\}$, $M := \max\{\mathcal{U}_2^0, \mathcal{U}_4^0, \mathcal{W}_2^0, \mathcal{W}_4^0\}$, and

$$w^0 := (M, M) \in \mathbb{R}_+^2 \quad \text{and} \quad u^0 := (m, m) \in \mathbb{R}_+^2. \tag{2.1}$$

We shall use the uniform coupling of all four chains $\{\mathcal{U}^t\}$, $\{\mathcal{W}^t\}$, $\{u^t\}$, and $\{w^t\}$. Observing that $u^0 \preceq \{\mathcal{U}^0, \mathcal{W}^0\} \preceq w^0$, we see that the uniform coupling keeps $\mathcal{U}^t$ and $\mathcal{W}^t$ perpetually "squeezed" between $u^t$ and $w^t$ (i.e., $u^t \preceq \{\mathcal{U}^t, \mathcal{W}^t\} \preceq w^t$ for all $t$). Corollary 2.1 below justifies why it suffices to consider the coupled pair $(u^t, w^t)$ in order to bound $d_{\mathrm{TV}}(\mathcal{U}^t, \mathcal{W}^t)$. But first we need a lemma.

**Lemma 2.1.** *Suppose that* $0 < \beta_1 < \beta_2 < \beta_3 < \beta_4$ *and* $\alpha > 0$. *Let* $f_i$ *be the density function of* $Z_i \sim \Gamma(\alpha, \beta_i)$. *Then*

$$\min\{f_1(y), f_4(y)\} \le \min\{f_2(y), f_3(y)\} \qquad \textit{for all } y \ge 0.$$

***Remark 2.1.*** Since a property of total variation (see Proposition 3 of [14]) is that

$$d_{\mathrm{TV}}(Z_i, Z_j) = 1 - \int \min\{f_i(y), f_j(y)\}\, dy,$$

we also conclude from Lemma 2.1 that $d_{\mathrm{TV}}(Z_2, Z_3) \le d_{\mathrm{TV}}(Z_1, Z_4)$.

**Proof of Lemma 2.1.** Note first that for $i, j \in \{1, 2, 3, 4\}$ with $i < j$,

$$f_i(y) \ge f_j(y) \iff \beta_i^\alpha \exp(-\beta_i y) \ge \beta_j^\alpha \exp(-\beta_j y) \iff y \ge g(\beta_i, \beta_j), \tag{2.2}$$

where

$$g(\beta_i, \beta_j) := \frac{\alpha(\ln(\beta_i) - \ln(\beta_j))}{\beta_i - \beta_j}.$$

Observe that $g(\beta_i, \beta_j)$ is the slope of the secant line joining the points $(\beta_i, z_i)$ and $(\beta_j, z_j)$ on the curve $z = \alpha \ln \beta$. Since this curve is concave, the slope is decreasing in $\beta_i$ and $\beta_j$ for $\beta_i < \beta_j$ (Lemma 5.16 of [16]), which implies

$$g(\beta_4, \beta_3) \le g(\beta_4, \beta_2) \le g(\beta_4, \beta_1) = g(\beta_1, \beta_4) \le g(\beta_1, \beta_3) \le g(\beta_1, \beta_2). \tag{2.3}$$

Then from (2.2) and (2.3) it follows that

$$f_1(y) \le \min\{f_2(y), f_3(y)\} \qquad \text{on } [0, g(\beta_1, \beta_3)] \quad \text{and}$$
$$f_4(y) \le \min\{f_2(y), f_3(y)\} \qquad \text{on } [g(\beta_4, \beta_2), \infty);$$

hence $\min\{f_1(y), f_4(y)\} \le \min\{f_2(y), f_3(y)\}$ on $[0, g(\beta_1, \beta_3)] \cup [g(\beta_4, \beta_2), \infty) = [0, \infty)$. □

We now describe "one-shot coupling" of the Markov chains $u$, $w$, $\mathcal{U}$, and $\mathcal{W}$ at time $t+1$ (described in [13] in greater generality). Assume that the uniform coupling of these chains holds up to and including time $t$. We shall also use the same two random variables $\gamma_1^{t+1}$ and $\gamma_3^{t+1}$ for all four chains. The following algorithm for generating $u^{t+1}$, $w^{t+1}$, $\mathcal{U}^{t+1}$, and $\mathcal{W}^{t+1}$ implicitly constructs different (dependent) values of $\gamma_2^{t+1}$ and $\gamma_4^{t+1}$ for each chain.

For $i \in \{2, 4\}$, let $f_{u_i}$ be the probability density function of the conditional distribution of $u_i^{t+1}$ given $u^t$ and $(\gamma_1^{t+1}, \gamma_3^{t+1})$, with analogous definitions for $f_{w_i}$, $f_{\mathcal{U}_i}$, and $f_{\mathcal{W}_i}$. For each coordinate $i \in \{2, 4\}$, we take $u_i^{[t+1]C}$ to be the $x$-coordinate of a uniformly chosen point from the area under the graph of the density function $f_{u_i}$. (The superscript $[t+1]C$ denotes that the coupling occurs at time $t+1$.) If this point also lies below the graph of the density function $f_{w_i}$, then set $w_i^{[t+1]C} = \mathcal{W}_i^{[t+1]C} = \mathcal{U}_i^{[t+1]C} = u_i^{[t+1]C}$. Otherwise, let $w_i^{[t+1]C}$ be the $x$-coordinate of a uniformly and independently chosen point from the area above the graph of $\min\{f_{u_i}, f_{w_i}\}$ and below the graph of $f_{w_i}$ (in this case, $w^{[t+1]C} \ne u^{[t+1]C}$ because $f_{u_i}(u^{[t+1]C}) > f_{w_i}(u^{[t+1]C})$), let $\mathcal{W}_i^{[t+1]C}$ be the $x$-coordinate of a uniformly and independently chosen point from the area above the graph of $\min\{f_{u_i}, f_{w_i}\}$ and below the graph of $f_{\mathcal{W}_i}$, and let $\mathcal{U}_i^{[t+1]C}$ be the $x$-coordinate of a uniformly and independently chosen point from the area above the graph of $\min\{f_{u_i}, f_{w_i}\}$ and below the graph of $f_{\mathcal{U}_i}$. By Lemma 2.1, we know $\min\{f_{u_i}, f_{w_i}\} \le \min\{f_{\mathcal{W}_i}, f_{\mathcal{U}_i}\}$, hence it is easy to verify that $(\mathcal{U}^{[t+1]C}, \mathcal{W}^{[t+1]C}, u^{[t+1]C}, w^{[t+1]C})$ is indeed a coupling of $\mathcal{U}^{t+1}$, $\mathcal{W}^{t+1}$, $u^{t+1}$, and $w^{t+1}$. (Observe that the relations $u^{[t+1]C} \preceq \{\mathcal{U}^{[t+1]C}, \mathcal{W}^{[t+1]C}\} \preceq w^{[t+1]C}$ may not hold.)

**Corollary 2.1.** *For one-shot coupling at time $t+1$, we have*

$$d_{\mathrm{TV}}(\mathcal{U}^{t+1}, \mathcal{W}^{t+1}) \le \mathbb{P}[u^{[t+1]C} \ne w^{[t+1]C}].$$

**Proof.** By the coupling construction, $\{\mathcal{U}_i^{[t+1]C} \ne \mathcal{W}_i^{[t+1]C}\} \subseteq \{u_i^{[t+1]C} \ne w_i^{[t+1]C}\}$ for $i = 2, 4$. Therefore

$$d_{\mathrm{TV}}(\mathcal{U}^{t+1}, \mathcal{W}^{t+1}) \le \mathbb{P}[\mathcal{U}^{[t+1]C} \ne \mathcal{W}^{[t+1]C}] \le \mathbb{P}[u^{[t+1]C} \ne w^{[t+1]C}]. \qquad \square$$

***Remark 2.2.*** It was shown at the beginning of this section that it suffices to couple initial points which satisfy the partial order "$\preceq$", and that the *uniform* coupling preserves this order. It is however also easy to verify that under this coupling alone, $\mathbb{P}[u^t \ne w^t] = 1$ whenever $u^0 \ne w^0$. For this reason, we introduced the *one-shot* coupling, for which $\mathbb{P}[u^{[t+1]C} \ne w^{[t+1]C}]$ can be small.

# 3. The ratio $R_t$

In this section, we continue to assume that $\mathcal{U}^0, \mathcal{W}^0 \in \mathbb{R}_+^2$ are two arbitrary starting points of two chains $\mathcal{U}$ and $\mathcal{W}$, that $u^0$ and $w^0$ are defined as in (2.1), and that all four chains $\mathcal{U}$, $\mathcal{W}$, $u$ and $w$ are constructed by the uniform coupling. Then $u^t \preceq \{\mathcal{U}^t, \mathcal{W}^t\} \preceq w^t$ and $w_i^t/u_i^t \geq 1$ for all $t \geq 0$ and $i \in \{2, 4\}$. In this section and the next, we shall mostly forget about the original chains $\mathcal{U}^t$ and $\mathcal{W}^t$, and focus the chains $u^t$ and $w^t$ with the goal of applying Corollary 2.1.

Define the filtration $\mathcal{F}_t := \sigma(u^0, w^0, \gamma_1^1, \ldots, \gamma_4^1, \ldots, \gamma_1^t, \ldots, \gamma_4^t)$. The following coupling construction will be used to define the non-increasing $\mathcal{F}_t$-measurable process $R_t$, with the property $R_t \geq \max_i\{\frac{w_i^t}{u_i^t}\}$. Note then that $u^t = w^t$ if $R_t = 1$.

Given our two initial points $u^0$ and $w^0$ (which satisfy $u^0 \preceq w^0$ by definition), we shall define two auxiliary processes $\tilde{v}$ and $v$. Let $v^0 := w^0$, so that by (2.1), $\frac{u_2^0}{u_4^0} = \frac{v_2^0}{v_4^0}$. Let $R_0 := \frac{v_2^0}{u_2^0}$ $(= \frac{v_4^0}{u_4^0})$. For each $t \geq 0$, we already have (recall (1.6))

$$\begin{aligned} u^{t+1} &= \left(u_2^{t+1}, u_4^{t+1}\right) := F^{t+1}\left(u_2^t, u_4^t\right) \\ &= \left(\frac{\gamma_2^{t+1}}{\gamma_1^{t+1}/(x+u_2^t) + \gamma_3^{t+1}/(u_2^t+u_4^t)}, \frac{\gamma_4^{t+1}}{\gamma_3^{t+1}/(u_2^t+u_4^t)+b}\right). \end{aligned}$$

For each $t \geq 0$, we recursively define

$$\begin{aligned} \tilde{v}^{t+1} &= \left(\tilde{v}_2^{t+1}, \tilde{v}_4^{t+1}\right) := F^{t+1}\left(v_2^t, v_4^t\right) \\ &= \left(\frac{\gamma_2^{t+1}}{\gamma_1^{t+1}/(x+v_2^t) + \gamma_3^{t+1}/(v_2^t+v_4^t)}, \frac{\gamma_4^{t+1}}{\gamma_3^{t+1}/(v_2^t+v_4^t)+b}\right), \\ R_{t+1} &:= \max\left\{\frac{\tilde{v}_2^{t+1}}{u_2^{t+1}}, \frac{\tilde{v}_4^{t+1}}{u_4^{t+1}}\right\}, \quad \text{and} \\ v^{t+1} &= \left(v_2^{t+1}, v_4^{t+1}\right) := \left(R_{t+1}u_2^{t+1}, R_{t+1}u_4^{t+1}\right). \end{aligned} \tag{3.1}$$

Note that unlike $u^t$, the process $v^t$ is not a Markov chain. Observe also that equality of ratios is preserved: $\frac{u_2^{t+1}}{u_4^{t+1}} = \frac{v_2^{t+1}}{v_4^{t+1}}$, and $\frac{v_2^{t+1}}{u_2^{t+1}} = \frac{v_4^{t+1}}{u_4^{t+1}} = R_{t+1}$. As mentioned in Section 1.2, the process $R_t$ serves as a statistic on the ratio-wise proximity of $u^t$ to $v^t$. It will follow that $w^t \preceq v^t$ for all $t$, thus $R_t$ is also an upper bound on the ratio-wise proximity of $u^t$ to $w^t$ (and hence also of the two chains $\mathcal{U}^t$ and $\mathcal{W}^t$). It will also follow from Lemma 3.2 that $R_t$ is non-increasing, which is not the case if we try to replace $\tilde{v}^{t+1}$ by $w^{t+1}$ in its definition. This issue is behind our motivation for the definition of $R_t$.

Recall that $w^0 = v^0$ and $w^{t+1} = F^{t+1}(w_t)$ for $t \geq 0$. Then by induction, the monotonicity of $F$ guarantees that $u^t \preceq w^t \preceq \tilde{v}^t \preceq v^t$ for every $t$. That is, the process $v^t$ dominates a copy of the Markov chain started at $w^0$ and coupled uniformly with $u^t$.

Before deriving properties of $R_t$, we state the following elementary calculus lemma.

**Lemma 3.1.** *Suppose that* $0 < a < b$. *Then* $g(x, y) := (\frac{x}{b} + y)/(\frac{x}{a} + y)$ *is decreasing in* $x$ *and increasing in* $y$, *for all* $x, y > 0$.

We shall now show that $\{R_t\}$ is *non-increasing*. Let

$$Q_t := \max\left\{\frac{\gamma_3^{t+1} + bu_4^t}{\gamma_3^{t+1} + bv_4^t}, \frac{\gamma_3^{t+1} + \gamma_1^{t+1}/(1 + x/u_2^t)}{\gamma_3^{t+1} + \gamma_1^{t+1}/(1 + x/v_2^t)}\right\}.$$

**Lemma 3.2.** *We have* $R_{t+1} \le Q_t R_t$ *and* $Q_t \le 1$ *for all* $t \ge 0$.

**Proof.** Since $u^t \preceq v^t$, it is immediate that $Q_t \le 1$. And by Lemma 3.1, we have

$$\begin{aligned}
R_{t+1} &= \max\left\{\frac{\gamma_3^{t+1}/(u_2^t + u_4^t) + b}{\gamma_3^{t+1}/(v_2^t + v_4^t) + b}, \frac{\gamma_3^{t+1}/(u_2^t + u_4^t) + \gamma_1^{t+1}/(u_2^t + x)}{\gamma_3^{t+1}/(v_2^t + v_4^t) + \gamma_1^{t+1}/(v_2^t + x)}\right\} \\
&= \frac{v_2^t}{u_2^t} \cdot \max\left\{\left(\frac{\gamma_3^{t+1}/(u_2^t/u_4^t + 1) + bu_4^t}{\gamma_3^{t+1}/(u_2^t/u_4^t + 1) + bv_4^t}\right), \right. \\
&\qquad \left.\left(\frac{\gamma_3^{t+1}/(u_4^t/u_2^t + 1) + \gamma_1^{t+1}/(1 + x/u_2^t)}{\gamma_3^{t+1}/(u_4^t/u_2^t + 1) + \gamma_1^{t+1}/(1 + x/v_2^t)}\right)\right\} \\
&\le R_t Q_t.
\end{aligned} \tag{3.2}$$

□

Lemma 3.2 shows that the sequence $\{R_t\}$ is non-increasing and that

$$\mathbb{E}[R_{t+1}] \le R_0 \mathbb{E}\left[\prod_{j=0}^{t} Q_j\right].$$

The next lemma shows that $\mathbb{P}[u^{[t+1]C} \neq w^{[t+1]C} | \mathcal{F}_t]$ is small if $R_t$ is close to 1.

**Lemma 3.3.** *For one-shot coupling at time* $t + 1$, *we have*

$$\mathbb{P}\big[u^{[t+1]C} \neq w^{[t+1]C} | \mathcal{F}_t\big] \le 1 - R_t^{-(a_2+a_3+a_4+a_5)}.$$

**Proof.** For $i \in \{2, 4\}$ and $\mathcal{G}_t := \sigma(\mathcal{F}_t, \gamma_1^{t+1}, \gamma_3^{t+1})$, let $f_{u_i}(y)$ and $f_{w_i}(y)$ be the conditional density functions of $u_i^{t+1}$ and $w_i^{t+1}$ given $\mathcal{G}_t$, as in our description of one-shot coupling in Section 2. By (1.6), these are gamma densities with shape parameters $a_i + a_{i+1}$, and inverse scale parameters $\Delta_{u,2}^{t+1} := \frac{\gamma_1^{t+1}}{x+u_2^t} + \frac{\gamma_3^{t+1}}{u_2^t+u_4^t}$ and $\Delta_{u,4}^{t+1} := b + \frac{\gamma_3^{t+1}}{u_2^t+u_4^t}$, with $\Delta_{w,2}^{t+1}$ and $\Delta_{w,4}^{t+1}$ defined similarly. Observe that $\Delta_{u,i}^{t+1} \ge \Delta_{w,i}^{t+1}$. Then for all $y > 0$,

$$f_{w_i}(y) \ge \left(\frac{\Delta_{w,i}^{t+1}}{\Delta_{u,i}^{t+1}}\right)^{a_i+a_{i+1}} f_{u_i}(y)$$

and therefore

$$\min\left\{f_{u_i}(y), f_{w_i}(y)\right\} \geq \left(\frac{\Delta_{w,i}^{t+1}}{\Delta_{u,i}^{t+1}}\right)^{a_i+a_{i+1}} f_{u_i}(y).$$

Recall that $v_4^t/v_2^t = u_4^t/u_2^t$, and observe that

$$\begin{aligned}\frac{\Delta_{u,2}^{t+1}}{\Delta_{w,2}^{t+1}} &\leq \frac{\gamma_1^{t+1}/(x+u_2^t)+\gamma_3^{t+1}/(u_2^t+u_4^t)}{\gamma_1^{t+1}/(x+v_2^t)+\gamma_3^{t+1}/(v_2^t+v_4^t)} \\ &= \frac{v_2^t}{u_2^t} \times \frac{\gamma_1^{t+1}/(x/u_2^t+1)+\gamma_3^{t+1}/(1+u_4^t/u_2^t)}{\gamma_1^{t+1}/(x/v_2^t+1)+\gamma_3^{t+1}/(1+v_4^t/v_2^t)} \leq \frac{v_2^t}{u_2^t} = R_t\end{aligned}$$

with a similar inequality following for $\Delta_{u,4}^{t+1}/\Delta_{w,4}^{t+1}$. By our construction of the one-shot coupling,

$$\begin{aligned}\mathbb{P}\left[u_i^{[t+1]C} \neq w_i^{[t+1]C} | \mathcal{G}_t\right] &= 1 - \int \min\left\{f_{u_i}(y), f_{w_i}(y)\right\} dy \\ &\leq 1 - \int f_{u_i}(y)\left(\frac{\Delta_{w,i}^{t+1}}{\Delta_{u,i}^{t+1}}\right)^{a_i+a_{i+1}} dy \\ &\leq 1 - R_t^{-a_i-a_{i+1}} \int f_{u_i}(y)\, dy \\ &= 1 - R_t^{-a_i-a_{i+1}}.\end{aligned}$$

Since the final bound is independent of $(\gamma_1^{t+1}, \gamma_3^{t+1})$, we also get $\mathbb{P}[u_i^{[t+1]C} \neq w_i^{[t+1]C} | \mathcal{F}_t] \leq 1 - R_t^{-a_i-a_{i+1}}$. Therefore,

$$\begin{aligned}\mathbb{P}\left[u^{[t+1]C} \neq w^{[t+1]C} | \mathcal{F}_t\right] &= \mathbb{P}\left[\bigcup_i \left\{u_i^{[t+1]C} \neq w_i^{[t+1]C}\right\} | \mathcal{F}_t\right] \\ &= 1 - \prod_{i=2,4} \mathbb{P}\left[\left\{u_i^{[t+1]C} = w_i^{[t+1]C}\right\} | \mathcal{F}_t\right] \\ &\leq 1 - R_t^{-a_2-a_3} R_t^{-a_4-a_5}.\end{aligned}$$

□

As we have seen, our ratio $R_t$ satisfies $R_t \geq \max\{\frac{w_i^t}{u_i^t}\}$, which is the condition stated at the beginning of Section 3. Our aim now is to show that $R_t$ converges to 1 at a geometric rate, or more explicitly to obtain an expression of the form

$$\mathbb{E}[R_{t+1}] \leq 1 + C_{R_0} \prod_{j=1}^{t+1} r_j,$$

where $r_j < 1$ and $r_j$ is "frequently" bounded from above by some $r < 1$ (the exact meaning of this will become apparent following the definition of $\bar{S}_t$ in (4.4)). Note that in order to achieve this, it suffices to have for all $t \geq 0$

$$\mathbb{E}[Q_t R_t] \leq r_{t+1}\big(\mathbb{E}[R_t] - 1\big) + 1. \tag{3.3}$$

Recall that $\mathcal{F}_t := \sigma(u^0, w^0, \gamma_1^1, \dots, \gamma_4^1, \dots, \gamma_1^t, \dots, \gamma_4^t)$. The left-hand side of (3.3) can be written as

$$\mathbb{E}[Q_t R_t] = \mathbb{E}\big[R_t \mathbb{E}[Q_t | \mathcal{F}_t]\big], \tag{3.4}$$

and we approximate $\mathbb{E}[Q_t|\mathcal{F}_t]$ in the following lemma.

**Lemma 3.4.** *Let* $\mu_1 = \mathbb{E}[\gamma_3] = a_3 + a_4$ *and* $\mu_2 = \mathbb{E}[\gamma_1 - \frac{1}{3}] = a_1 + a_2 - \frac{1}{3}$, *and let* $\hat{r}_t = 1 - 1/\max\{(\frac{4\mu_1}{\mu_2} + 4)(\frac{u_2^t}{x} + \frac{x}{v_2^t} + 2), 4 + \frac{4\mu_1}{bv_4^t}\}$. *If $S$ is a $\mathcal{F}_t$-measurable stopping time, then*

$$\mathbb{E}[Q_S|\mathcal{F}_S] \leq \hat{r}_S + \frac{1 - \hat{r}_S}{R_S}.$$

*Thus we have*

$$\mathbb{E}[Q_S R_S] \leq \mathbb{E}\big[\hat{r}_S(R_S - 1)\big] + 1.$$

**Proof.** By [2], we have $\mathbb{P}[\gamma_3 \leq \mu_1] \geq \frac{1}{2}$ and $\mathbb{P}[\gamma_1 \geq \mu_2] \geq \frac{1}{2}$. Hence by Lemma 3.1, for any $t$, the probability is at least $\frac{1}{4}$ that $Q_t \leq \max\{(\frac{\mu_1 + \mu_2/(1 + x/u_2^t)}{\mu_1 + \mu_2/(1 + x/v_2^t)}), (\frac{\mu_1 + bu_4^t}{\mu_1 + bv_4^t})\}$. Then, using $R_S = v_2^S/u_2^S = v_4^S/u_4^S \geq 1$, we have

$$\begin{aligned}
\mathbb{E}[Q_S|\mathcal{F}_S] &\leq \frac{1}{4} \cdot \max\left\{\left(\frac{\mu_1 + \mu_2/(1 + x/u_2^S)}{\mu_1 + \mu_2/(1 + x/v_2^S)}\right), \left(\frac{\mu_1 + bu_4^S}{\mu_1 + bv_4^S}\right)\right\} + 1 \cdot \frac{3}{4} \\
&= \frac{1}{4} \cdot \max\left\{\left(1 - \frac{1/(1 + x/v_2^S) - 1/(1 + x/u_2^S)}{\mu_1/\mu_2 + 1/(1 + x/v_2^S)}\right), \left(1 - \frac{bv_4^S - bu_4^S}{\mu_1 + bv_4^S}\right)\right\} + \frac{3}{4} \\
&= 1 - \frac{(1 - 1/R_S)}{4\max\{(\mu_1/\mu_2 + 1/(1 + x/v_2^S))(u_2^S/x + 1)(x/v_2^S + 1), 1 + \mu_1/(bv_4^S)\}} \qquad (3.5) \\
&\leq 1 - \frac{(1 - 1/R_S)}{4\max\{(\mu_1/\mu_2 + 1)(u_2^S/x + x/v_2^S + 2), 1 + \mu_1/(bv_4^S)\}} \\
&= \hat{r}_S + \frac{1 - \hat{r}_S}{R_S}.
\end{aligned}$$

□

Our task in the next section will be to show that we frequently have $\hat{r}_t \leq r$ for some $r < 1$, which by Lemma 3.4 will result in an expression of the form given by (3.3).

# 4. Auxiliary processes with drift conditions

We continue to work with the processes described at the beginning of Section 3. We first define random processes $K_{1,t}$ and $K_{2,t}$ ($t \geq 0$) and constants $\zeta_i$ and $C_i$ as follows.

$$K_{1,t} := u_2^t + u_4^t, \qquad K_{2,t} := \begin{cases} \dfrac{u_3^t + u_1^t + b}{\gamma_2^t + \gamma_4^t}, & \text{if } t \geq 1, \\ \dfrac{1}{u_2^0 + u_4^0}, & \text{if } t = 0, \end{cases}$$

$$\zeta_1 := \frac{a_2 + a_3}{a_1 + a_2 + a_3 + a_4 - 1}, \qquad \zeta_2 := \frac{a_3 + a_4}{a_2 + a_3 + a_4 + a_5 - 1},$$

$$C_1 := \zeta_1 x + \frac{a_4 + a_5}{b}, \qquad C_2 := \frac{a_1 + a_2 + xb}{x(a_2 + a_3 + a_4 + a_5 - 1)}.$$

Both $K_{1,t}$ and $K_{2,t}$ are adapted to $\mathcal{F}_t$ and are in fact functions of $\vec{u}^t$ (since $\gamma_2^t + \gamma_4^t = u_2^t(u_1^t + u_3^t) + u_4^t(u_3^t + b)$ for $t \geq 1$). The following inequalities ("drift conditions") are important for us.

The proofs of all lemmas in this section are given in Section 4.1.

**Lemma 4.1.** *For $i = 1, 2$ and $t \geq 0$, we have*

$$\mathbb{E}[K_{i,t+1} | \mathcal{F}_t] \leq \zeta_i K_{i,t} + C_i. \tag{4.1}$$

Note that condition (1.7) guarantees that $\zeta_1 < 1$ and $\zeta_2 < 1$. We shall assume that condition (1.7) holds for the rest of this section.

Next, we define some more processes for $t \geq 1$, all adapted to $\mathcal{F}_t$:

$$\rho_t := \max\left\{ \left(\frac{4\mu_1}{\mu_2} + 4\right)\left(\frac{u_2^t}{x} + \frac{x}{v_2^t} + 2\right), 4 + \frac{4\mu_1}{b v_4^t} \right\},$$

$$D_t := \frac{1}{x}\left(\frac{4\mu_1}{\mu_2} + 4\right)(u_2^t + u_4^t) + \left(\left(\frac{4\mu_1}{\mu_2} + 4\right)x + \frac{4\mu_1}{b}\right)\left(\frac{1}{u_2^t} + \frac{1}{u_4^t}\right),$$

$$\omega_{1,t} := \frac{1}{x}\left(\frac{4\mu_1}{\mu_2} + 4\right)\frac{\gamma_2^t}{\gamma_1^t + \gamma_3^t}, \qquad \tilde{\omega}_{2,t} = 2 + \frac{\gamma_2^t}{\gamma_4^t} + \frac{\gamma_4^t}{\gamma_2^t},$$

$$\omega_{2,t} := \left(\left(\frac{4\mu_1}{\mu_2} + 4\right)x + \frac{4\mu_1}{b}\right)\tilde{\omega}_{2,t}\frac{\gamma_3^t}{\gamma_2^t + \gamma_4^t},$$

$$\omega_{3,t} := \frac{1}{x}\left(\frac{4\mu_1}{\mu_2} + 4\right)\left(\frac{\gamma_2^t}{\gamma_1^t + \gamma_3^t}x + \frac{\gamma_4^t}{b}\right) + \left(\left(\frac{4\mu_1}{\mu_2} + 4\right)x + \frac{4\mu_1}{b}\right)\tilde{\omega}_{2,t}\frac{\gamma_1^t/x + b}{\gamma_2^t + \gamma_4^t}.$$

Observe that $\hat{r}_t = 1 - 1/\rho_t$. Clearly, $(\omega_{1,t}, \omega_{2,t}, \omega_{3,t})$ is a nonnegative random vector that is i.i.d. over time $t \geq 1$, measurable with respect to $\mathcal{F}_t$ and independent of $\mathcal{F}_{t-1}$.

Our interest in these processes comes from the following. As explained at the end of the previous section, we want to show that the random process $\hat{r}_t$ is frequently less than some fixed $r < 1$. Writing $\hat{r}_t = 1 - 1/\rho_t$, this is equivalent to showing that $\rho_t$ is frequently less than some fixed bound. The inequalities in the following lemma provide us with alternative quantities to bound in order to achieve our goal.

**Lemma 4.2.** *For all* $t \geq 1$, *we have*

$$\rho_t \leq D_t \quad \text{and} \quad D_{t+1} \leq \omega_{1,t+1} K_{1,t} + \omega_{2,t+1} K_{2,t} + \omega_{3,t+1}. \tag{4.2}$$

To help bound the rightmost expression in equation (4.2), we define the process

$$J_t := K_{1,t} + K_{2,t} \qquad \text{for } t \geq 0$$

and the constants

$$A := \max\{\zeta_1, \zeta_2\}, \qquad C := C_1 + C_2, \qquad \eta := \frac{2C}{1-A}, \qquad \beta := \frac{1}{2}(1+A).$$

By condition (1.7), we have that $A < 1$ and $\beta < 1$. It is easy to see that

$$\mathbb{E}[J_{t+1} | \mathcal{F}_t] \leq A J_t + C \qquad \text{for all } t \geq 0,$$

from which one can directly calculate (observe $A\eta + C = \beta\eta$) that

$$\mathbb{E}[J_{t+1} | \mathcal{F}_t] \leq \begin{Bmatrix} \beta J_t, & \text{if } J_t \geq \eta \\ \beta\eta, & \text{if } J_t \leq \eta \end{Bmatrix} = \beta \max\{J_t, \eta\}. \tag{4.3}$$

The following lemma will be used to show that $J$ is frequently below $\eta$. This will be used with Lemma 4.2 to help bound $D$, and hence $\rho$.

**Lemma 4.3.** *For* $s \geq 1$, *let* $L_s = \{i \in [1, s] | J_i \geq \eta\}$. *Then for every* $s \geq 1$,

(a) $\mathbb{P}[L_s = \mathcal{L} | \mathcal{F}_0] \leq \beta^{|\mathcal{L}|} \max\{J_0, \eta\}/\eta$ *for every* $\mathcal{L} \subseteq [1, s]$, *and*
(b) *For every* $\mathcal{L} \subseteq [1, s]$ *with* $s \in \mathcal{L}$, *we have* $\mathbb{E}[J_s \mathbf{1}_{\{L_s = \mathcal{L}\}} | \mathcal{F}_0] \leq \beta^{|\mathcal{L}|} \max\{J_0, \eta\}$.

The term $\beta^{|\mathcal{L}|}$ in Lemma 4.3 indicates that $L_s$ is unlikely to be large.

The next lemma uses the relation $\hat{r}_t \leq 1 - 1/D_t$. The idea is that if $J_t$ is bounded, then $D_t$ is probably not too large, and $\hat{r}_t$ is not too close to 1.

**Lemma 4.4.** *Let* $S \geq 1$ *be an a.s. finite stopping time adapted to* $\mathcal{F}_t$ *such that* $J_S < \eta$, *and let* $0 \leq Y \in \mathcal{F}_S$. *Define* $\theta_i := \mathbb{E}[\omega_{i,t+1}]$ *for* $i = 1, 2, 3$, *and let*

$$r = 1 - \frac{1}{(\theta_1 + \theta_2)\eta + \theta_3}.$$

*Then* $\mathbb{E}[Y(R_{S+2} - 1)] \leq r\mathbb{E}[Y(R_S - 1)]$. *In particular,* $\mathbb{E}[R_{S+2} - 1] \leq r\mathbb{E}[R_S - 1]$.

For the following important result, we write

$$\bar{S}_t := [1,t] \setminus L_t = \{i \in [1,t] | J_i < \eta\} \qquad \text{for } t \geq 1. \tag{4.4}$$

By applying Lemma 4.4 to those times that are in $\bar{S}_t$, we obtain the following.

**Lemma 4.5.** *In the event* $\{J_0 < \eta\}$, *we have*

$$\mathbb{E}\left[(R_{t+2} - 1)\mathbf{1}_{|\bar{S}_t|>k} | \mathcal{F}_0\right] \leq r^{\lceil (k+1)/2 \rceil}(R_0 - 1).$$

The above lemma leads us to the following theorem, which is an explicit decay rate for the ratio $R_t$. Its proof is in Section 4.1. The idea is that by Lemma 4.3, $|\bar{S}_t|$ is unlikely to be small. Indeed, we show that it is unlikely to be smaller than roughly $t/d$. Hence the approximate fraction of time that $J_s < \eta$ (and hence $\hat{r}_s < r$) is at least $1/d$.

**Theorem 4.1.** *Let* $d = \max\{3, 2\ln(\beta|\ln\beta|\sqrt{r}/2)/\ln\beta\}$. *Then in* $\{J_0 < \eta\}$, *we have* $\mathbb{E}[R_{t+2}|\mathcal{F}_0] \leq 1 + 3r^{t/2d}(R_0 - 1)$ *for all* $t > 0$.

Theorem 4.1 is the last main ingredient that we need to prove Theorem 1.1.

**Proof of Theorem 1.1.** It will be convenient here to perform the "one-shot coupling" at time $t+3$ rather than at time $t+1$. By Corollary 2.1, $\mathbb{P}[u^{[t+3]C} \neq w^{[t+3]C}]$ is an upper bound for $d_{\mathrm{TV}}(\mathcal{U}^{t+3}, \mathcal{W}^{t+3})$. First, we restrict to the event $\{J_0 < \eta\}$. Theorem 4.1 tells us that

$$\mathbb{E}[R_{t+2} - 1 | \mathcal{F}_0] \leq 3r^{t/2d}(R_0 - 1).$$

Therefore by Lemma 3.3, Jensen's inequality, and the bound $1 - (1+y)^{-p} \leq py$ for $p, y \geq 0$ (easily shown by calculus),

$$\begin{aligned}
\mathbb{P}\left[u^{[t+3]C} \neq w^{[t+3]C} | \mathcal{F}_0\right] &= \mathbb{E}\left[\mathbb{P}\left[u^{[t+3]C} \neq w^{t+3} | \mathcal{F}_{t+2}\right] | \mathcal{F}_0\right] \\
&\leq \mathbb{E}\left[1 - (R_{t+2})^{-(a_2+a_3+a_4+a_5)}\right] \leq 1 - \left(\mathbb{E}[R_{t+2}]\right)^{-(a_2+a_3+a_4+a_5)} \\
&\leq 1 - \left(1 + 3r^{t/2d}(R_0 - 1)\right)^{-(a_2+a_3+a_4+a_5)} \\
&\leq 3r^{t/2d}(a_2+a_3+a_4+a_5)(R_0 - 1).
\end{aligned}$$

This proves the first statement of the theorem. If we no longer restrict to the event $\{J_0 < \eta\}$, then let $T = \min\{t \geq 0 : J_t < \eta\}$. Recalling Lemma 4.3, we see that on the event $\{J_0 \geq \eta\}$ we have $T > t_0$ if and only if $L_{t_0} = [1, t_0]$. Therefore, on $\{J_0 \geq \eta\}$ we have

$$\begin{aligned}
&\mathbb{P}\left[u^{[t+3]C} \neq w^{[t+3]C} | \mathcal{F}_0\right] \\
&\quad \leq \mathbb{P}\left[u^{[t+3]C} \neq w^{[t+3]C} | \mathcal{F}_0, T \leq \left\lfloor \frac{t}{2} \right\rfloor + 3\right] + \mathbb{P}\left[T > \left\lfloor \frac{t}{2} \right\rfloor + 3 \middle| \mathcal{F}_0\right] \\
&\quad \leq 3r^{t/4d}(a_2+a_3+a_4+a_5)(R_0 - 1) + \frac{\max\{J_0, \eta\}\beta^{\lfloor t/2 \rfloor + 3}}{\eta}.
\end{aligned} \tag{4.5}$$

Since this is greater than what we have on $\{J_0 < \eta\}$, it is also a bound for general values of $J_0$. □

## 4.1. Remaining proofs

**Proof of Lemma 4.1.** Observe that for $t \geq 0$

$$\begin{aligned} u_2^{t+1} + u_4^{t+1} &= \frac{\gamma_2^{t+1}}{\gamma_1^{t+1}/(x+u_2^t) + \gamma_3^{t+1}/(u_2^t+u_4^t)} + \frac{\gamma_4^{t+1}}{\gamma_3^{t+1}/(u_2^t+u_4^t)+b} \\ &\leq \frac{\gamma_2^{t+1}}{\gamma_1^{t+1}+\gamma_3^{t+1}}\big(u_2^t+u_4^t+x\big) + \frac{\gamma_4^{t+1}}{b}. \end{aligned} \tag{4.6}$$

Therefore, $\mathbb{E}[K_{1,t+1}|\mathcal{F}_t] \leq \zeta_1 K_{1,t} + C_1$. Observe that since

$$u_3^{t+1} = \frac{\gamma_3^{t+1}}{u_2^t+u_4^t} = \frac{\gamma_3^{t+1}}{\gamma_2^t/(u_1^t+u_3^t)+\gamma_4^t/(u_3^t+b)} \leq \frac{\gamma_3^{t+1}}{\gamma_2^t+\gamma_4^t}\big(u_1^t+u_3^t+b\big) = \gamma_3^{t+1}K_{2,t}$$

for $t \geq 1$, it follows that

$$K_{2,t+1} \leq \frac{\gamma_3^{t+1}}{\gamma_2^{t+1}+\gamma_4^{t+1}}K_{2,t} + \frac{u_1^{t+1}+b}{\gamma_2^{t+1}+\gamma_4^{t+1}} \tag{4.7}$$

and hence

$$\mathbb{E}[K_{2,t+1}|\mathcal{F}_t] \leq \zeta_2 K_{2,t} + \mathbb{E}\left[\frac{\gamma_1^{t+1}/x+b}{\gamma_2^{t+1}+\gamma_4^{t+1}}\middle|\mathcal{F}_t\right] \leq \zeta_2 K_{2,t} + C_2 \tag{4.8}$$

for $t \geq 0$ (the $t = 0$ case is immediate from the definition of $K_{2,0}$). $\square$

**Proof of Lemma 4.2.** The inequality $\rho_t \leq D_t$ follows from the facts that $u^t \preceq v^t$ and $2 \leq \frac{u}{x} + \frac{x}{u}$. Note next that

$$\frac{1}{u_2^{t+1}} + \frac{1}{u_4^{t+1}} \leq \left(\frac{1}{\gamma_2^{t+1}} + \frac{1}{\gamma_4^{t+1}}\right)\big(u_1^{t+1}+u_3^{t+1}+b\big) = \tilde{\omega}_{2,t+1}K_{2,t+1} \tag{4.9}$$

and $\tilde{\omega}_{2,t+1}$ is independent of $\mathcal{F}_t$. By (4.6), (4.7) and (4.9) we conclude that for $t \geq 1$,

$$\begin{aligned} D_{t+1} &\leq \frac{1}{x}\left(\frac{4\mu_1}{\mu_2}+4\right)\left(\frac{\gamma_2^{t+1}}{\gamma_1^{t+1}+\gamma_3^{t+1}}(K_{1,t}+x) + \frac{\gamma_4^{t+1}}{b}\right) \\ &\quad + \left(\left(\frac{4\mu_1}{\mu_2}+4\right)x + \frac{4\mu_1}{b}\right)\tilde{\omega}_{2,t+1}\left(\frac{\gamma_3^{t+1}}{\gamma_2^{t+1}+\gamma_4^{t+1}}K_{2,t} + \frac{\gamma_1^{t+1}/x+b}{\gamma_2^{t+1}+\gamma_4^{t+1}}\right) \\ &\leq \omega_{1,t+1}K_{1,t} + \omega_{2,t+1}K_{2,t} + \omega_{3,t+1}. \end{aligned}$$

$\square$

**Proof of Lemma 4.3.** The proof will be by induction. For $s = 1, 2, \ldots,$ let $Sa(s)$ and $Sb(s)$ be the statements of parts (a) and (b) respectively, namely:

*Sa*($s$): "$\mathbb{P}[L_s = \mathcal{L}|\mathcal{F}_0] \le \beta^{|\mathcal{L}|}\max\{J_0,\eta\}/\eta$ for every $\mathcal{L} \subseteq [1,s]$";
*Sb*($s$): "For every $\mathcal{L} \subseteq [1,s]$ with $s \in \mathcal{L}$, we have $\mathbb{E}[J_s\mathbf{1}_{\{L_s=\mathcal{L}\}}|\mathcal{F}_0] \le \beta^{|\mathcal{L}|}\max\{J_0,\eta\}$."

We first verify the case $s = 1$. In *Sa*(1), the inequality is trivial if $\mathcal{L} = \varnothing$, while for $\mathcal{L} = \{1\}$, Markov's inequality and equation (4.3) show that

$$\mathbb{P}\big[L_1 = \{1\}|\mathcal{F}_0\big] = \mathbb{P}[J_1 \ge \eta|\mathcal{F}_0] \le \frac{\mathbb{E}[J_1|\mathcal{F}_0]}{\eta} \le \frac{\beta\max\{J_0,\eta\}}{\eta}.$$

This verifies *Sa*(1). For *Sb*(1), we only need to consider $\mathcal{L} = \{1\}$, for which

$$\mathbb{E}[J_1\mathbf{1}_{\{L_1=\{1\}\}}|\mathcal{F}_0] \le \mathbb{E}[J_1|\mathcal{F}_0] \le \beta\max\{J_0,\eta\}$$

by equation (4.3). This verifies *Sb*(1).

Now assume that *Sa*($t$) and *Sb*($t$) hold, and we shall deduce *Sa*($t+1$) and *Sb*($t+1$). Let $\mathcal{L} \subset [1, t+1]$, and let $\mathcal{L}' = \mathcal{L} \setminus \{t+1\}$. We consider cases, according as to whether or not $\mathcal{L}$ contains $t+1$ and $t$.

*Case* I: $t+1 \notin \mathcal{L}$. In this case, $\mathcal{L} = \mathcal{L}'$, so by *Sa*($t$) we have

$$\mathbb{P}[L_{t+1} = \mathcal{L}|\mathcal{F}_0] \le \mathbb{P}\big[L_t = \mathcal{L}'|\mathcal{F}_0\big] \le \frac{\beta^{|\mathcal{L}'|}\max\{J_0,\eta\}}{\eta} = \frac{\beta^{|\mathcal{L}|}\max\{J_0,\eta\}}{\eta}.$$

This verifies the inequality in (a). There is nothing to check for (b).

*Case* II: $t+1 \in \mathcal{L}$. First, we have

$$\mathbb{E}[J_{t+1}\mathbf{1}_{\{L_{t+1}=\mathcal{L}\}}|\mathcal{F}_0] \le \mathbb{E}\big(\mathbb{E}[J_{t+1}|\mathcal{F}_t]\mathbf{1}_{\{L_t=\mathcal{L}'\}}|\mathcal{F}_0\big). \tag{4.10}$$

We now consider subcases II(i) and II(ii), based on $t$.

*Case* II(i): $t+1 \in \mathcal{L}$ and $t \notin \mathcal{L}$. Since $\{L_t = \mathcal{L}'\} \subset \{J_t < \eta\}$, we obtain from equation (4.3) that $\mathbb{E}[J_{t+1}|\mathcal{F}_t]\mathbf{1}_{\{L_t=\mathcal{L}'\}} \le \beta\eta\mathbf{1}_{\{L_t=\mathcal{L}'\}}$. Hence by equation (4.10) and *Sa*($t$), we obtain

$$\begin{aligned}\mathbb{E}[J_{t+1}\mathbf{1}_{\{L_{t+1}=\mathcal{L}\}}|\mathcal{F}_0] \le \beta\eta\mathbb{E}[\mathbf{1}_{\{L_t=\mathcal{L}'\}}|\mathcal{F}_0] &\le \beta\eta\frac{\beta^{|\mathcal{L}'|}\max\{J_0,\eta\}}{\eta} \\ &= \beta^{|\mathcal{L}|}\max\{J_0,\eta\}.\end{aligned}$$

This proves the inequality in *Sb*($t+1$). This in turn implies

$$\mathbb{P}[L_{t+1} = \mathcal{L}|\mathcal{F}_0] \le \frac{\mathbb{E}[J_{t+1}\mathbf{1}_{\{L_{t+1}=\mathcal{L}\}}|\mathcal{F}_0]}{\eta} \le \frac{\beta^{|\mathcal{L}|}\max\{J_0,\eta\}}{\eta}.$$

Thus the inquality in *Sa*($t+1$) also holds for $\mathcal{L}$.

*Case* II(ii): $t+1 \in \mathcal{L}$ and $t \in \mathcal{L}$. Since $\{L_t = \mathcal{L}'\} \subset \{J_t \ge \eta\}$, we obtain from equation (4.3) that $\mathbb{E}[J_{t+1}|\mathcal{F}_t]\mathbf{1}_{\{L_t=\mathcal{L}'\}} \le \beta J_t\mathbf{1}_{\{L_t=\mathcal{L}'\}}$. Hence by equation (4.10) and *Sb*($t$) [note that $t \in \mathcal{L}'$], we obtain

$$\begin{aligned}\mathbb{E}[J_{t+1}\mathbf{1}_{\{L_{t+1}=\mathcal{L}\}}|\mathcal{F}_0] \le \beta\mathbb{E}[J_t\mathbf{1}_{\{L_t=\mathcal{L}'\}}|\mathcal{F}_0] &\le \beta\beta^{|\mathcal{L}'|}\max\{J_0,\eta\} \\ &= \beta^{|\mathcal{L}|}\max\{J_0,\eta\}.\end{aligned}$$

This proves the inequality in $Sb(t+1)$. The inequality for $Sa(t+1)$ now follows as in Case II(i). $\square$

**Proof of Lemma 4.4.** We start by observing that $D_{S+1} \le \eta(\omega_{1,S+1} + \omega_{2,S+1}) + \omega_{3,S+1}$. Therefore, applying Lemma 3.4 we get

$$\begin{aligned}
\mathbb{E}\big[Y(R_{S+2})\big] &\le \mathbb{E}[Y Q_{S+1} R_{S+1}] \\
&\le \mathbb{E}\big[Y\hat{r}_{S+1}(R_{S+1}-1)\big] + \mathbb{E}[Y] \qquad (4.11)\\
&\le \mathbb{E}\left[Y\left(1-\frac{1}{D_{S+1}}\right)(R_S-1)\right] + \mathbb{E}[Y] \qquad (\text{using } R_{t+1} \le R_t)\\
&\le \mathbb{E}\left[Y\left(1-\frac{1}{\eta(\omega_{1,S+1}+\omega_{2,S+1})+\omega_{3,S+1}}\right)(R_S-1)\right] + \mathbb{E}[Y]\\
&= \mathbb{E}\left[\left(1-\frac{1}{\eta(\omega_{1,S+1}+\omega_{2,S+1})+\omega_{3,S+1}}\right)\right]\mathbb{E}\big[Y(R_S-1)\big] + \mathbb{E}[Y]\\
&\le r\mathbb{E}\big[Y(R_S-1)\big] + \mathbb{E}[Y] \qquad (\text{by Jensen's inequality}). \qquad (4.12)
\end{aligned}$$

$\square$

**Proof of Lemma 4.5.** Let $\tau_0 = 0$ and $\{\tau_i\} \subseteq \{1, 2, \ldots\}$ be those times for which $J_{\tau_i} < \eta$. Then by Lemma 4.4 with $Y = \mathbf{1}_{\tau_{k+1} \le t}$ and $S = \tau_{k+1}$,

$$\begin{aligned}
\mathbb{E}[R_{t+2}\mathbf{1}_{|\bar{S}_t|>k}|\mathcal{F}_0] &= \mathbb{E}[R_{t+2}\mathbf{1}_{\tau_{k+1}\le t}|\mathcal{F}_0]\\
&\le \mathbb{E}[R_{\tau_{k+1}+2}\mathbf{1}_{\tau_{k+1}\le t}|\mathcal{F}_0]\\
&\le r\mathbb{E}\big[\mathbf{1}_{\tau_{k+1}\le t}(R_{\tau_{k+1}}-1)|\mathcal{F}_0\big] + \mathbb{P}\big[|\bar{S}_t| > k|\mathcal{F}_0\big]\\
&\le r\mathbb{E}\big[\mathbf{1}_{\tau_{k-1}\le t}(R_{\tau_{k-1}+2}-1)|\mathcal{F}_0\big] + \mathbb{P}\big[|\bar{S}_t| > k|\mathcal{F}_0\big].
\end{aligned}$$

The last inequality uses the fact that $\mathbf{1}_{\tau_{k+1}\le t} \le \mathbf{1}_{\tau_{k-1}\le t}$ and $R_{\tau_{k+1}} \le R_{\tau_{k-1}+2}$. This then leads to the first step in an inductive argument:

$$\begin{aligned}
&\mathbb{E}[R_{\tau_{k+1}+2}\mathbf{1}_{\tau_{k+1}\le t}|\mathcal{F}_0] - \mathbb{P}\big[|\bar{S}_t| > k|\mathcal{F}_0\big]\\
&\quad \le r\big(\mathbb{E}[R_{\tau_{k-1}+2}\mathbf{1}_{\tau_{k-1}\le t}|\mathcal{F}_0] - \mathbb{P}\big[|\bar{S}_t| > k-2|\mathcal{F}_0\big]\big).
\end{aligned} \qquad (4.13)$$

Proceeding in this manner, we claim that we get

$$\mathbb{E}[R_{\tau_{k+1}+2}\mathbf{1}_{\tau_{k+1}\le t}|\mathcal{F}_0] - \mathbb{P}\big[|\bar{S}_t| > k|\mathcal{F}_0\big] \le r^{\lceil (k+1)/2\rceil}(R_0-1).$$

The ceiling function in the exponent $\lceil (k+1)/2\rceil$ is immediate whenever $k+1$ is even. If on the other hand $k+1$ is odd, then by (4.13) and Lemma 4.4, we have

$$\begin{aligned}
\mathbb{E}[R_{\tau_{k+1}+2}\mathbf{1}_{\tau_{k+1}\le t}|\mathcal{F}_0] - \mathbb{P}\big[|\bar{S}_t| > k|\mathcal{F}_0\big] &\le r^{\lfloor (k+1)/2\rfloor}\mathbb{E}\big[\mathbf{1}_{\tau_1\le t}(R_{\tau_1+2}-1)|\mathcal{F}_0\big]\\
&\le r^{\lfloor (k+1)/2\rfloor} r\mathbb{E}\big[\mathbf{1}_{\tau_1\le t}(R_{\tau_1}-1)|\mathcal{F}_0\big]\\
&\le r^{\lfloor (k+1)/2\rfloor+1}(R_0-1).
\end{aligned}$$

$\square$

**Proof of Theorem 4.1.** For any $k < t$, we deduce from Lemmas 4.5 and 4.3(a) that

$$\begin{aligned}
\mathbb{E}[R_{t+2}|\mathcal{F}_0] &= \mathbb{E}[R_{t+2}\mathbf{1}_{|\bar{S}_t|>k}|\mathcal{F}_0] + \mathbb{E}[R_{t+2}\mathbf{1}_{|\bar{S}_t|\le k}|\mathcal{F}_0] \\
&\le r^{\lceil (k+1)/2\rceil}(R_0-1) + \mathbb{P}\big[|\bar{S}_t|>k|\mathcal{F}_0\big] + \mathbb{E}[R_0\mathbf{1}_{|\bar{S}_t|\le k}|\mathcal{F}_0] \\
&\le r^{\lceil (k+1)/2\rceil}(R_0-1) + \mathbb{P}\big[|\bar{S}_t|>k|\mathcal{F}_0\big] \\
&\quad +(R_0-1)\mathbb{P}\big[|\bar{S}_t|\le k|\mathcal{F}_0\big] + \mathbb{P}\big[|\bar{S}_t|\le k|\mathcal{F}_0\big] \\
&\le 1 + (R_0-1)\left(r^{\lceil (k+1)/2\rceil} + \sum_{j=0}^{k}\binom{t}{j}\beta^{t-j}\right).
\end{aligned} \tag{4.14}$$

Henceforth, let $k = \lfloor \frac{t}{d} \rfloor$. Since $k \le t/3$, we have $\binom{t}{j} \le \frac{1}{2}\binom{t}{j+1}$ for $j < k$ and hence $\sum_{j=0}^{k}\binom{t}{j}\beta^{t-j} \le 2\binom{t}{k}\beta^{t-k}$. Next, note that $\binom{t}{k}q^k(1-q)^{t-k} \le 1$ whenever $0 < q < 1$. Taking $q = 1/d$, we get

$$\begin{aligned}
\binom{t}{k} &\le d^k\left(1-\frac{1}{d}\right)^{-(t-k)} = \frac{d^t}{(d-1)^{t-k}} \le \frac{d^t}{(d-1)^{t-t/d}} \\
&= \left[d\left(1+\frac{1}{d-1}\right)^{d-1}\right]^{t/d} < (de)^{t/d}.
\end{aligned} \tag{4.15}$$

By calculus, we have $y\beta^y \le 2\beta^{y/2}/(e|\ln\beta|)$ for all $y > 0$. Combining this with results of the preceding paragraph, we obtain

$$\begin{aligned}
r^{\lceil (k+1)/2\rceil} + \sum_{j=0}^{k}\binom{t}{j}\beta^{t-j} &\le r^{(k+1)/2} + 2\binom{t}{k}\beta^{t-k} \\
&\le r^{t/2d} + 2\big(ed\beta^{d-1}\big)^{t/d} \\
&\le r^{t/2d} + 2\left(\frac{2}{\beta|\ln\beta|}\beta^{d/2}\right)^{t/d} \\
&\le 3r^{t/2d}.
\end{aligned}$$

Together with (4.14), this proves the desired bound. □

## 5. Sampling from equilibrium

It is not hard to apply our previous results to obtain a bound on the rate of convergence to the equilibrium distribution $\pi$ of the chain (1.6).

**Proof of Corollary 1.1.** Fix $\mathcal{U}^0$ and let $\mathcal{W}^0$ be a random vector with density $\pi$. Define $u^t$ and $w^t$ accordingly. By (4.5), we have

$$\mathbb{P}\big[u^{[t+3]C} \neq w^{[t+3]C} | \mathcal{W}_0\big] \leq 3r^{t/4d}(a_2+a_3+a_4+a_5)(R_0-1) + \frac{\max\{J_0,\eta\}\beta^{\lfloor t/2\rfloor+3}}{\eta}.$$

The corollary now follows from Corollary 2.1. □

Now let $C_g := \int(\prod_{i=1}^4 z_i^{a_i+a_{i+1}-1})\exp(\sum_{i=1}^5 -z_i z_{i-1})\,dz$. Then we can bound the terms $\mathbb{E}_\pi[R_0]$ and $\mathbb{E}_\pi[J_0]$ in Corollary 1.1 in the following way:

$$\begin{aligned} d_{\mathrm{TV}}\big(\mathcal{U}^{t+3},\pi\big) \leq{}& 3r^{t/4d}(a_2+a_3+a_4+a_5)\frac{1}{C_g} \\ & \times \int\left(\frac{\max\{1,v_2,v_4\}}{\min\{1,v_2,v_4\}}\right)\left(\prod_{i=1}^4 v_i^{a_i+a_{i+1}-1}\right)\exp\left(\sum_{i=1}^5 -v_i v_{i-1}\right)dv \\ & + \frac{\beta^{\lfloor t/2\rfloor+3}}{\eta}\left(\eta + \frac{1}{C_g}\int J_0\left(\prod_{i=1}^4 v_i^{a_i+a_{i+1}-1}\right)\exp\left(\sum_{i=1}^5 -v_i v_{i-1}\right)dv\right) \\ \leq{}& 3\tilde{C_\pi} r^{t/4d}(a_2+a_3+a_4+a_5) + \left(\frac{\tilde{C_J}}{\eta}+1\right)\beta^{\lfloor t/2\rfloor+3}, \end{aligned}$$

where $\tilde{C_\pi} := \int(\frac{\max\{1,v_2,v_4\}}{\min\{1,v_2,v_4\}})(\prod_{i=1}^4 v_i^{a_i+a_{i+1}-1})\exp(\sum_{i=1}^5 -v_i v_{i-1})\,dv/C_g$ and $\tilde{C_J} := \int J_0(\prod_{i=1}^4 v_i^{a_i+a_{i+1}-1})\exp(\sum_{i=1}^5 -v_i v_{i-1})\,dv/C_g$. We derive bounds for these terms in Appendix B in [10].

For the purpose of illustrating this result in a concrete example, let us set $x=2$, $b=3$ and $a_i=i$. From Appendix B in [10] we get $\tilde{C_\pi} \leq 31\,065$ $\tilde{C_J} \leq 59$, $\beta \leq 7/9$, $r \leq 1-\frac{3}{4356}$, $10 \leq \eta \leq 11$ and $18 \leq d \leq 20$. Hence

$$d_{\mathrm{TV}}\big(\mathcal{U}^{t+3},\pi\big) \leq 31\,065 * 43\left(1-\frac{3}{4356}\right)^{t/80} + \left(1+\frac{59}{20}\right)\left(\frac{7}{9}\right)^{\lfloor t/2\rfloor+3}$$

which implies that $d_{\mathrm{TV}}(\mathcal{U}^{t+3},\pi) \leq 10^{-5}$ for $t \geq 2\,100\,000$.

# 6. A brief look at the case $n=3$

The case $n=3$ can be treated in a very similar manner as was used for $n=4$. The problem reduces to dealing with a Markov chain of a single variable, namely the second coordinate of the three, given by

$$u^{t+1} = \frac{\gamma_2^{t+1}}{\gamma_1^{t+1}/(u^t+x) + \gamma_3^{t+1}/(u^t+b)}. \tag{6.1}$$

The uniform coupling of two chains $u^t$ and $w^t$ with the property $u^0 \leq w^0$ results in $u^t \leq v^t$ for all $t$. If $u^0 < w^0$, then it is not hard to see that the ratio $R_t = \frac{w^t}{u^t}$ is strictly decreasing, hence we no longer need to define a process like (3.1) and we can simply work with this ratio directly. Indeed, $R_{t+1} = R_t Q_t$ where

$$\begin{aligned} Q_t &:= 1 - \frac{(1-1/R_t)(x\gamma_1^{t+1}/((u^t+x)(1+x/w^t)) + b\gamma_3^{t+1}/((u^t+b)(1+b/w^t)))}{(\gamma_1^{t+1}/(1+x/w^t) + \gamma_3^{t+1}/(1+b/w^t))} \\ &\leq 1 - \frac{(1-1/R_t)(x\gamma_1^{t+1} + b\gamma_3^{t+1})}{(\gamma_1^{t+1}/(1+x/w^t) + \gamma_3^{t+1}/(1+b/w^t))(u^t + \max\{x,b\})(1+\max\{x,b\}/u^t)} \\ &\leq \hat{r}_t + \frac{1-\hat{r}_t}{R_t}, \end{aligned}$$

where $\hat{r}_t := 1 - \min\{x,b\}/((u^t + \max\{x,b\})(1 + \frac{\max\{x,b\}}{u^t}))$. Note that if we define $K_{1,t+1} := u^{t+1}$ and $K_{2,t+1} := \frac{1}{u^{t+1}}$ then $K_{1,t+1} \leq \frac{\gamma_2^{t+1}}{\gamma_1^{t+1}+\gamma_3^{t+1}}(u^t + x + b)$ and $K_{2,t+1} \leq (\frac{\gamma_1^{t+1}}{\gamma_2^{t+1}}\frac{1}{x} + \frac{\gamma_3^{t+1}}{\gamma_2^{t+1}}\frac{1}{b})$, and hence we do not need a process analogous to $D_t$ from Section 4, since

$$\hat{r}_{t+1} \leq 1 - \min\{x,b\}/\big(\big(K_{1,t+1} + \max\{x,b\}\big)\big(1 + \max\{x,b\}K_{2,t+1}\big)\big).$$

As before, we will require that $a_1 + a_4 > 1$ in order that $\mathbb{E}[\gamma_2/(\gamma_1+\gamma_3)] < 1$, and $a_2 + a_3 > 1$ in order that $\mathbb{E}[\frac{\gamma_1^{t+1}}{\gamma_2^{t+1}}] < \infty$. If $J_t := K_{1,t} + K_{2,t}$ and $S$ is a measurable stopping time such that $J_S \leq \eta$, with

$$\eta := 2\left(\frac{(x+b)(a_2+a_3)}{a_1+a_2+a_3+a_4-1} + \frac{(a_1+a_2)/x + (a_3+a_4)/b}{a_2+a_3-1}\right)\bigg/\left(1 - \frac{a_2+a_3}{a_1+a_2+a_3+a_4-1}\right)$$

then we can repeat the steps of the proof of Lemma 4.4 as follows:

$$\begin{aligned} \mathbb{E}[R_{S+1}] &= \mathbb{E}[Q_S R_S] \\ &\leq \mathbb{E}\big[\hat{r}_S(R_S - 1)\big] + 1 \\ &= \mathbb{E}\left[\left(1 - \frac{\min\{x,b\}}{(u^S + \max\{x,b\})(1+\max\{x,b\}/u^S)}\right)(R_S - 1)\right] + 1 \\ &\leq \mathbb{E}\left[\left(1 - \frac{\min\{x,b\}}{(u^S + 2\max\{x,b\} + \max\{x,b\}^2/u^S)}\right)(R_S - 1)\right] \\ &\leq \mathbb{E}\left[\left(1 - \frac{\min\{x,b\}}{(2\max\{x,b\} + \eta(1+\max\{x,b\}^2))}\right)(R_S - 1)\right] + 1 \\ &= r\mathbb{E}[R_S - 1] + 1, \end{aligned} \tag{6.2}$$

where $r = 1 - \min\{x,b\}/(2\max\{x,b\} + \eta(1 + \max\{x,b\}^2))$. Note that we no longer need to look at time $S+2$ in the left-hand side of (6.2) in order to obtain this inequality. This means that from the proof of Lemma 4.5 and Corollary 4.1 we get

$$\mathbb{E}[R_{t+1}|J_0 \le \eta] \le 1 + 3r^{t/d}(R_0 - 1),$$

where $d = \max\{3, 2\ln(\beta|\ln\beta|r/2)/\ln\beta\}$. From the proof of Theorem 1.1 we conclude

**Theorem 6.1 [$n = 3$].** *Suppose that $a_1 + a_4 > 1$ and $a_2 + a_3 > 1$. If $u^t$ and $w^t$ are two instances of the Markov chain (6.1), then*

$$d_{\mathrm{TV}}\big(u^{t+2}, w^{t+2}\big) \le r^{t/(2d)}\big(1 + 3(a_2 + a_3)(R_0 - 1)\big) + \frac{\max\{J_0, \eta\}\beta^{\lfloor t/2\rfloor+3}}{\eta}.$$

We can make an analogous argument to obtain a result similar to Corollary 1.1. In particular, if we let $\mathcal{U}^0 = (1, 1, 1)$, $\mathcal{W}^0 \sim \pi$ and $x = 1$, $b = 2$ and $a_i = i$, then by calculations similar to those done in Section 5 we get

$$d_{\mathrm{TV}}\big(\mathcal{U}^{t+2}, \pi\big) \le 600\left(\frac{78}{79}\right)^{t/40} + 6\left(\frac{7}{9}\right)^{\lfloor t/2\rfloor+3}$$

which in particular implies that $d_{\mathrm{TV}}(\mathcal{U}^{t+2}, \pi) \le 10^{-5}$ for $t \ge 50\,000$.

# Appendix

$$C_1 = \tfrac{a_2+a_3}{a_1+a_2+a_3+a_4-1}x + \tfrac{a_4+a_5}{b},\ C_2 = \tfrac{a_1+a_2+xb}{x(a_2+a_3+a_4+a_5-1)}$$

$$\varrho = \tfrac{4(a_3+a_4)}{(a_1+a_2-1/3)}$$

$$\eta = \tfrac{C_1+C_2}{1-\max\{(a_2+a_3)/(a_1+a_2+a_3+a_4-1),(a_3+a_4)/(a_2+a_3+a_4+a_5-1)\}}$$

$$\theta_1 = \tfrac{1}{x}(\varrho+4)\tfrac{a_2+a_3}{a_1+a_2+a_3+a_4-1}$$

$$\theta_2 = \mathbb{E}[(2 + \tfrac{\gamma_2}{\gamma_4} + \tfrac{\gamma_4}{\gamma_2})(\tfrac{\gamma_3}{\gamma_2+\gamma_4})]((\varrho+4)x + \tfrac{4(a_3+a_4)}{b})$$

$$\theta_3 = \tfrac{1}{x}(\varrho+4)(\tfrac{a_2+a_3}{a_1+a_2+a_3+a_4-1}x + \tfrac{a_4+a_5}{b}) + ((\varrho+4)x + \tfrac{a_4+a_5}{b})\mathbb{E}[(2 + \tfrac{\gamma_2}{\gamma_4} + \tfrac{\gamma_4}{\gamma_2})(\tfrac{\gamma_1/x+b}{\gamma_2+\gamma_4})]$$

$$r = 1 - (\eta(\theta_1+\theta_2)+\theta_3)^{-1}$$

$$\beta = \tfrac{1+\max\{(a_2+a_3)/(a_1+a_2+a_3+a_4-1),(a_3+a_4)/(a_2+a_3+a_4+a_5-1)\}}{2}$$

$$d = \max\{3, 2\ln(\beta|\ln\beta|\sqrt{r}/2)/\ln\beta\}$$

We can calculate $\theta_2$ and $\theta_3$ with the help of partial fractions, as follows. Writing $A_i = a_i + a_{i+1}$, we obtain

$$\mathbb{E}\left(\left(\frac{\gamma_2}{\gamma_4} + \frac{\gamma_4}{\gamma_2}\right)\frac{1}{\gamma_2+\gamma_4}\right) = \mathbb{E}\left(\frac{1}{\gamma_2} + \frac{1}{\gamma_4} - \frac{2}{\gamma_2+\gamma_4}\right) = \frac{A_2^2 + A_4^2 - A_2 - A_4}{(A_2-1)(A_4-1)(A_2+A_4-1)}.$$

## Acknowledgements

Oliver Jovanovski is supported through NWO Gravitation Grant 024.002.003-NETWORKS. The research of Neal Madras is supported in part by a Discovery Grant from the Natural Science and Engineering Research Council of Canada. The authors thank the referees for their comments and suggestions, which have helped make this paper much more readable.

## References

[1] Besag, J., Green, P., Higdon, D. and Mengersen, K. (1995). Bayesian computation and stochastic systems. *Statist. Sci.* **10** 3–66. MR1349818
[2] Chen, J. and Rubin, H. (1986). Bounds for the difference between median and mean of gamma and Poisson distributions. *Statist. Probab. Lett.* **4** 281–283. MR0858317
[3] Constantinou, A., Fenton, N. and Neil, M. (2012). pi-football: A Bayesian network model for forecasting association football match outcomes. *Knowl.-Based Syst.* **36** 322–339.
[4] Cowans, P. (2004). Information retrieval using hierarchical Dirichlet process. In *Proceedings of the Annual International Conference on Research and Development in Information Retrieval* **27** 564–565. New York: ACM.
[5] Diaconis, P. and Freedman, D. (1999). Iterated random functions. *SIAM Rev.* **41** 45–76. MR1669737
[6] Díez, F.J., Mira, J., Iturralde, E. and Zubillaga, S. (1997). DIAVAL a Bayesian expert system for echocardiography. *Artif. Intell. Med.* **10** 59–73.
[7] Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.* **6** 721–741.
[8] Jiang, X. and Cooper, G.F. A Bayesian spatio-temporal method for disease outbreak detection. *J. Am. Med. Inform. Assoc.* **17** 462–471.
[9] Jiang, X., Neapolitan, R., Barmada, M. and Visweswaran, S. (2011). Learning genetic epistasis using Bayesian network scoring criteria. *BMC Bioinformatics* **12** 89.
[10] Jovanovski, O. and Madras, N. (2014). Convergence rates for hierarchical Gibbs samplers. Available at arXiv:1402.4733v1.
[11] Lindvall, T. (1992). *Lectures on the Coupling Method. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*. New York: Wiley. MR1180522
[12] Madras, N. and Sezer, D. (2010). Quantitative bounds for Markov chain convergence: Wasserstein and total variation distances. *Bernoulli* **16** 882–908. MR2730652
[13] Roberts, G.O. and Rosenthal, J.S. (2002). One-shot coupling for certain stochastic recursive sequences. *Stochastic Process. Appl.* **99** 195–208. MR1901153
[14] Roberts, G.O. and Rosenthal, J.S. (2004). General state space Markov chains and MCMC algorithms. *Probab. Surv.* **1** 20–71. MR2095565

[15] Rossi, P.E., Allenby, G.M. and McCulloch, R. (2005). *Bayesian Statistics and Marketing*. *Wiley Series in Probability and Statistics*. Chichester: Wiley. MR2193403
[16] Royden, H.L. (1988). *Real Analysis*, 3rd ed. New York: Macmillan Publishing Company. MR1013117
[17] Stephens, M., Smith, N. and Donnelly, P. (2001). A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* **68** 978–989.
[18] Tierney, L. (1994). Markov chains for exploring posterior distributions. *Ann. Statist.* **22** 1701–1762. MR1329166

*Received October 2014 and revised July 2015*