

CONVERGENCE RATE ANALYSIS OF MARKOV CHAINS

OLIVER JOVANOVSKI

A DISSERTATION SUBMITTED TO THE FACULTY OF GRADUATE
STUDIES
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN DEPARTMENT OF MATHEMATICS AND
STATISTICS
YORK UNIVERSITY
TORONTO, ONTARIO
MAY 2014

©Oliver Jovanovski, 2014

Abstract

We consider a number of Markov chains and derive bounds for the rate at which convergence to equilibrium occurs. For our main problem, we establish results for the rate of convergence in total variation of a Gibbs sampler to its equilibrium distribution. This sampler is motivated by a hierarchical Bayesian inference construction for a gamma random variable. The Bayesian hierarchical method involves statistical models that incorporate prior beliefs about the likelihood of observed data to arrive at posterior interpretations, and appears in applications for information technology, statistical genetics, market research and others. Our results apply to a wide range of parameter values in the case that the hierarchical depth is 3 or 4, and are more restrictive for depth greater than 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 co-ordinates. We construct auxiliary stochastic processes to show that this ratio does converge to 1 at a geometric rate. In addition, we also consider a stochastic image restoration model proposed by A.

Gibbs, and give an upper bound on the time it takes for a Markov chain defined by this model to be arbitrarily close in total variation to equilibrium. We use Gibbs' result for convergence in the Wasserstein metric to arrive at our result. Our bound for the time to equilibrium is of similar order to that of Gibbs.

Acknowledgements

I would like to express my gratitude to my advisor, Neal Madras, for all his guidance, support and unwavering patience.

Table of Contents

Abstract	ii
Acknowledgements	iv
Table of Contents	v
List of Figures	viii
1 Introduction	1
1.1 Markov chains and convergence	1
1.2 MCMC and Gibbs algorithms	3
1.3 Some known results	6
1.4 Our work	10
2 Hierarchical Gibbs Sampler	15
2.1 Random dynamical systems	16
2.2 Multiple data points	27

3	An attempt at generalisation	33
4	A different approach	38
4.1	The problem	38
4.2	Outline of our proof	44
4.3	Stochastically monotone coupling	45
4.4	The ratio R_t	49
4.5	Special case $n = 4$	52
4.6	Super-martingale-type auxiliary processes	56
4.7	Construction of D_t	67
4.8	Sampling from equilibrium	71
4.9	A brief look at the case $n = 3$	73
4.10	$n > 4$	76
4.11	Construction of $\{K_{i,t}\}$ for $n \geq 5$	84
4.12	Simulated results	92
5	An Image Restoration Model	96
5.1	Introduction	96
5.2	From d_W to d_{TV}	100
	Bibliography	107

Appendices	109
Appendix A	109
Appendix B	115

List of Figures

4.1	Simulated convergence for $n = 3$ case	93
4.2	Simulated convergence for $n = 4$ case	94

1 Introduction

The aim of this introductory chapter will be to provide a brief background to Markov chains, their convergence properties and already established results and methods, as well as our contribution to this subject. We will also give a preview of some of the methods used by us to accomplish these goals.

1.1 Markov chains and convergence

A Markov chain is a stochastic process $\{X_t\}_{t=0}^{\infty}$ characterized by its 'memoryless' property (also called the Markov property): given the current state of the chain, its future trajectory is independent of the past. This process takes place on some agreed-on set called the *state space*, and often denoted by Ω . The most common way to represent all of this is through a *probability kernel*, which is a map $P : \Omega \times \mathcal{S} \rightarrow [0, 1]$ (where \mathcal{S} is a sigma-algebra on Ω) satisfying

1. $P(x, \cdot)$ is a probability measure on Ω , for all $x \in \Omega$
2. $P(\cdot, A)$ is a \mathcal{S} -measurable function, for all $A \in \mathcal{S}$

In this context, $P(x, A)$ would simply represent the (transition) probability that $X_{t+1} \in A$, given $X_t = x$ (for a thorough discussion on this subject see [18]).

Under some fairly general conditions (see for example [1]), a Markov chain will converge (in distribution, as well as in the stronger measure of total variation, defined in equation (1.1)) to an *equilibrium probability distribution*, usually denoted by π . This property is exploited in a variety of Markov Chain Monte Carlo (MCMC) algorithms, which allow for the sampling from a distribution that is of arbitrary proximity to some specified *target distribution* (the probability distribution one would like to sample from, and hence also the equilibrium distribution of the Markov chain). This is particularly useful when the target distribution is too complicated to work with directly.

The utility of sampling from a distribution may be seen in approximating difficult integrals. For example, it may be desired to evaluate an integral of the form

$$\int f(x) d\pi = \mathbb{E}_\pi[f]$$

and in many cases, this could be a formidable task. By the law of large numbers this can be approximated with arbitrary degree of accuracy simply by taking a sufficient number of independent samples $\{s_1, \dots, s_k\}$ from the distribution π and evaluating $\frac{1}{k} \sum_{i=1}^k f(s_i)$.

1.2 MCMC and Gibbs algorithms

An MCMC algorithm involves the construction of a Markov chain which has as its equilibrium the targeted probability distribution, and to which it hence converges (a survey of some common algorithms can be found in [13] and [19]). However, mere knowledge of convergence is often not enough and it may be of both theoretical and practical interest to consider the rate at which convergence transpires (and while in finite state spaces convergence is always at a geometric rate (see for example Chapter 3 of [7]), this is not the case in a generalised state space). In particular, deriving an upper bound on the rate of convergence would provide a rigorously defined degree of certainty to how far this Markov chain is from its equilibrium distribution, and it would allow to assess the efficiency of an algorithm. Here the 'efficiency' of a Markov chain $\{X_t\}$ refers to how long it is necessary to run the Markov chain in order to obtain a given amount of data.

In this thesis we consider the problem of finding bounds for the rate at which certain Markov chains converge to equilibrium. To quantify this, we refer to two methods commonly used to measure the distance to equilibrium. These are by means of the *total variation metric* and *Wasserstein metric*. If ν_1 and ν_2 are two probability measures on the same state space Ω , then the total variation distance

between ν_1 and ν_2 is defined by

$$d_{TV}(\nu_1, \nu_2) := \sup_{A \subseteq \Omega} |\nu_1(A) - \nu_2(A)| \quad (1.1)$$

while their Wasserstein distance is defined by

$$d_W(\nu_1, \nu_2) := \inf \mathbb{E} [d(X, Y)] \quad (1.2)$$

where the infimum is taken over all joint distributions (X, Y) such that $X \sim \nu_1$ and $Y \sim \nu_2$, and $d(\cdot, \cdot)$ is an agreed upon metric on the state space. Furthermore, for two random variables $X \sim \nu_1$ and $Y \sim \nu_2$, we let $d_{TV}(X, Y) := d_{TV}(\nu_1, \nu_2)$ and $d_W(X, Y) := d_W(\nu_1, \nu_2)$.

It is not difficult to see that if the metric d on the state space is given by

$$d(x, y) := \begin{cases} 1 & x \neq y \\ 0 & x = y \end{cases}$$

then d_W is equal to d_{TV} . In general however, convergence in d_W does not necessarily imply convergence in d_{TV} and vice versa. For example, if we let Θ_n be equal to $\frac{1}{n}$ or 0 with probability $\frac{1}{2}$ and take the usual metric on \mathbb{R} , then $d_W(\Theta_n, 0) \rightarrow 0$. It is obvious, however, that $d_{TV}(\Theta_n, 0) = \frac{1}{2}$. Conversely, if Υ_n is equal to 0 or n with probability $1 - \frac{1}{n}$ and $\frac{1}{n}$ respectively, then $d_{TV}(\Upsilon_n, 0) \rightarrow 0$ while $d_W(\Upsilon_n, 0) = 1$ in the usual metric. For further examples, as well as some conditions that allow comparison between d_{TV} and d_W , see [4].

The Gibbs sampler [12] has been a very popular MCMC algorithm in obtaining a sample from a probability distribution that is otherwise difficult to work with. In its fundamental form, this algorithm works on a random 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 $(u_{j \neq i})$ (in other words, given u_j for all $j \neq i$). For example, let $h_i(w_i | u)$ be the conditional density function of u_i given $(u_{j \neq i})$. (Note that in this definition the variable u appearing in the condition is n -dimensional, however we only mean given the aforementioned $n - 1$ co-ordinates. We will use this convention throughout the thesis.) Then for $i \in \{1, 2, \dots, m\}$ we let

$$\bar{P}_i(v, dw) := \left(\prod_{j \neq i} \delta_{v_j}(w_j) \right) h_i(w_i | v) dw_i$$

In other words, \bar{P}_i is the probability kernel updating only the i^{th} co-ordinate, according to the conditional probability defined by $h_i(w_i | u)$. Then

$$\bar{P} := \frac{1}{m} \sum_{j=1}^m \bar{P}_j$$

is a probability kernel that represents uniformly choosing $i \in \{1, 2, \dots, m\}$ at every step in time, and updating that co-ordinate only. Alternatively, for any permutation $\sigma \in S_m$ we could define the probability kernel

$$Q := \prod \bar{P}_{\sigma(j)} \tag{1.3}$$

which updates the co-ordinates systematically, in the order $\sigma(1), \dots, \sigma(m)$. In

fact, as will be seen in Section 4, it is not necessary for the subscripts in the right-hand side of (1.3) to be a permutation and it could consist of more than m entries, so long as every term in $\{1, 2, \dots, m\}$ is included.

The main consideration in this paper is given to variants of the Gibbs sampler that converge to the equilibrium distribution of our model. In Section 4 we consider a few such variants chosen according to their advantages with respect to our problem. For $n = 4$ (dimensions of the vector, as discussed above) this choice can be characterised as a simultaneous updating algorithm. First we show that the problem can be simplified to a 2 dimensional Markov chain, and then at each step in time we set both components of u^{t+1} to be random with a distribution derived conditionally from u^t . For $n > 4$, we consider two different Markov chains given in (4.39) and (4.43). The evolution of both is elaborate, but we ultimately choose to work with (4.43) since it proves to be more favourable to our method.

1.3 Some known results

Numerous methods have been developed for analyzing the rate at which a Markov chain converges to equilibrium, including decomposition methods (e.g. [21],[20]), path coupling ([22]), comparison methods ([23]) and others. For Gibbs samplers in particular, some general convergence results have been derived (e.g. [14], [25], [24]), however due to their limitations it is often not possible to infer quantitative

bounds directly from these results. Another frequently used method to derive such bounds is to couple two copies of a Markov chain (for a detailed discussion on this subject see [2]), and evaluate the likelihood of coalescence at some future time. By this we mean the following: if $\{X_t\}$ and $\{Y_t\}$ are two Markov chains on the state space Ω , a (Markovian) *coupling* is a stochastic process $(\tilde{X}_t, \tilde{Y}_t)_{t=0}^\infty$ on Ω^2 such that $\{\tilde{X}_t\}_{t=0}^\infty$ is a Markov chain with the same law as $\{X_t\}_{t=0}^\infty$, and similarly for $\{\tilde{Y}_t\}_{t=0}^\infty$. *Coalescence* refers to the event $X_t = Y_t$. Note that once coalescence has been achieved at some time τ , we can redefine the coupling for times $\tau + s$, $s = 1, 2, \dots$ by setting $\tilde{Y}_{\tau+s} := \tilde{X}_{\tau+s}$. It is clear that this does indeed satisfy the definition of a coupling, since both $\{\tilde{X}_t\}$ and $\{\tilde{Y}_t\}$ remain faithful copies of the chain.

The following simple conclusion (known as the ‘‘Coupling Lemma’’) follows easily.

$$\begin{aligned}
& |\mathbb{P}[X_t \in A] - \mathbb{P}[Y_t \in A]| \\
&= \mathbb{P}[X_t \in A, X_t = Y_t] + \mathbb{P}[X_t \in A, X_t \neq Y_t] \\
&\quad - \mathbb{P}[Y_t \in A, X_t = Y_t] - \mathbb{P}[Y_t \in A, X_t \neq Y_t]| \\
&= |\mathbb{P}[X_t \in A, X_t \neq Y_t] - \mathbb{P}[Y_t \in A, X_t \neq Y_t]| \\
&\leq \mathbb{P}[X_t \neq Y_t]
\end{aligned}$$

Since this is true for any measurable set A , it follows that $d_{TV}(X_t, Y_t) \leq \mathbb{P}[X_t \neq Y_t]$.

It therefore suffices to determine how fast a coupling of two copies of a Markov chain achieves coalescence, in order to obtain a bound on the convergence rate of the chain. To illustrate this idea with an example, let us consider two simple symmetric random walks on a pentagon, commencing at points X_0 and Y_0 respectively. Both chains move clockwise or anticlockwise with probability $\frac{1}{4}$, and remain unchanged with probability $\frac{1}{2}$. We can couple X_t and Y_t by moving Y_t anticlockwise whenever X_t moves clockwise, and clockwise whenever X_t moves anticlockwise. We do this until they coalesce, and keep them together thereafter. Observe that regardless of the initial values X_0 and Y_0 , the probability of no coalescence within 2 steps is at most $1 - \frac{1}{16}$, therefore $d_{TV}(X_{2t}, \nu) \leq (1 - \frac{1}{16})^t$ where ν is the uniform probability distribution on the pentagon. It is shown in [1] that $d_{TV}(X_{2t}, \nu)$ is monotone decreasing in t , hence we can associate $\sqrt{1 - \frac{1}{16}}$ with a bound on the geometric rate at which this chain converges to equilibrium.

It may also be quite useful to seek out an appropriate partial order on the state space, and attempt to couple in a stochastically monotone manner that preserves this partial order for all time (a method employed in [3]). This has been part of our approach in Section 4: we define a partial order \preceq and consider initial vectors $u^0 \preceq \tilde{v}^0$. We define a coupling (u^t, \tilde{v}^t) that satisfies $u^t \preceq \tilde{v}^t$ for all time t , but rather than keeping track of these two Markov chains, we follow u^t in tandem with another stochastic process v^t that serves as a majorant (in the partial order \preceq) to

both copies of the Markov chain. We then provide an upper bound in Corollary 4.14 on the rate at which the ratio v^t/u^t converges geometrically to 1, which is ultimately used (in a manner similar to the “one-shot coupling” approach discussed in [10]) to obtain an upper bound on the rate at which the two chains $\{u^t\}$ and $\{\tilde{v}^t\}$ coalesce.

One further method of bounding the convergence rate of a Markov chain involves representing the chain as a trajectory of a random dynamical system (or iterated function system). In other words, if $\{v_t\}$ is the chain in question, then there is some sequence of i.i.d. random functions $\{f_i\}$ such that $v_t \sim f_t \left(f_{t-1} \left(\dots f_1(v_0) \dots \right) \right)$, where v_0 is the initial state. For example, we can define the i.i.d. random functions $\{f_i\}$ on \mathbb{R} by

$$f_i(v_0) = \frac{v_0}{1.1} + u_i$$

where u_i are i.i.d uniform random variables on some interval $[a, b]$. It is easy to see that the iterates

$$v_t := f_t \left(f_{t-1} \left(\dots f_1(v_0) \dots \right) \right) = \frac{v_0}{(1.1)^t} + \frac{u_1}{(1.1)^{t-1}} + \dots + u_t \quad (1.4)$$

define a Markov chain. With this notation in mind, we can now also consider the backwards iterated random system

$$\tilde{v}_t := f_1 \left(f_2 \left(\dots f_t(v_0) \dots \right) \right) = \frac{v_0}{(1.1)^t} + \frac{u_t}{(1.1)^{t-1}} + \dots + u_1 \quad (1.5)$$

which has the same distribution as v_t , but will usually not be Markov chain. However, it will in many cases converge pointwise to a random variable that has the

same distribution as the equilibrium distribution of v_t . For the example above, the pointwise convergence of (1.5) is quite obvious. This observation is used in [5] to show that for a random iterated function system $\{f_i\}$, if

$$\exists r < 1 \text{ s.t. } \forall z \limsup_{y \rightarrow z} \mathbb{E} \left[\frac{\rho(f_i(z), f_i(y))}{\rho(z, y)} \right] \leq r \quad (1.6)$$

where ρ is a metric on the state space, then \tilde{v}_t converges pointwise at a rate r (or faster), and hence v_t converges at the same rate (in the Wasserstein metric) to the aforementioned equilibrium. For our example (1.4), ρ is the Euclidean metric and $r = \frac{1}{1.1}$.

Definition 1.1. *A random iterated function system satisfying condition (1.6) shall be called 1-contractive.*

Note that this is a special case of a *locally contractive* iterated function system, as defined in [5].

1.4 Our work

Bayesian inference networks are popular statistical representations used to handle problems ranging from sports predictions and gambling to genetics, disease outbreak detection and artificial intelligence ([31], [32], [33], [34]). Gibbs samplers are frequently associated with problems in Bayesian inference, which is also the case for the problem considered in Section 2.1, Section 4 and Section 5.2, and motivated

by the Bayesian-inference scenario outlined in the beginning of Part 2. Similar constructions have appeared in numerous statistical models used in a variety of applications. In information retrieval related to search engines, hierarchical models are used to decide how to represent documents based on relevant queries (see for example [28]). Multi-Population Haplotype Phasing is a problem in statistical genetics where hierarchical Bayesian models can be used to represent genotypes (e.g. [29]), and in market research similar models are used in predicting buyer behaviour and decision making ([30]).

Our initial approach to this problem is in Section 2.1, focusing only on the case $n = 3$. Here the target distribution has the density function

$$f(u, v, w) \propto u^{a_1} e^{-xu} v^{a_2} u^{a_2-1} e^{-uv} w^{a_3} v^{a_3-1} e^{-vw} w^{a_4-1} e^{-wb} \quad (1.7)$$

where x is a known data point. We consider the Markov chain which sequentially updates u , v and w in that particular order, and show that the problem can be reduced to the one dimensional random dynamical system $\{f_t\}$ defined by

$$f_t(v) = \frac{\gamma_2^t}{\frac{\gamma_1^t}{x+v} + \frac{\gamma_3^t}{b+v}}$$

The randomness is attributed to the triplets of gamma random variables

$\left\{ (\gamma_1^t, \gamma_2^t, \gamma_3^t) \right\}_t$, described in more detail in Section 2.1. It is not hard to see that the local contractivity, as defined in (1.6), does not hold over the entire domain.

However, our observation is that the process \tilde{v}_t may nonetheless be 1-contractive

if viewed over a longer timespan. In other words, condition (1.6) may hold for the function $g(y) = f_1 \left(f_2 \left(\dots f_k(y) \dots \right) \right)$ and some fixed k , in which case \tilde{v}_t converges at a rate $\sqrt[k]{r}$ or faster. We show that for the hierarchical Gibbs sampler with $n = 3$

$$d_W(v_{kt+l}, \pi) \leq \frac{r^t}{1-r} \left(\frac{|x-b|}{1-\alpha} + x\alpha^l \right) \quad (1.8)$$

where k is defined in terms of the parameters $\{x, b, a_1, \dots, a_4\}$, $\alpha = \frac{a_2+a_3}{a_1+a_2+a_3+a_4-1}$ and $0 \leq l \leq k-1$.

In Section 2.2 we diverge somewhat from the aforementioned problem - we consider a similar Bayesian-inference setup with a hierarchical sampler of depth 2, but given multiple data points. Having used the same approach as in Section 2.1, our result is similar to the one stated in (1.8).

Section 3 is an attempt at finding a condition to generalise the method used in Section 2.1 to a wider range of random dynamical systems. We stipulate some conditions on real valued random functions $\{f_i\}$ which would imply that this random dynamical system converges in the long run, without being 1-contractive on the whole domain.

The method used in Section 2.1 does not readily translate to the case $n \geq 4$. To this end we develop a very different approach in Section 4. We consider two variations on a Gibbs sampler, both converging to π in the metric d_{TV} , and our objective becomes finding an upper bound on the distance between π and a copy of the chain with some initial state \mathcal{U}^0 . We use the first of the two versions (of a

Markov chain converging to π) primarily for the case $n = 4$. We show that if \mathcal{U}^t and \mathcal{V}^t are two instances of the chain and $d_{TV}(\mathcal{U}^t, \mathcal{V}^t)$ is the distance between their respective probability distributions at time t , then

$$d_{TV}(\mathcal{U}^{t+3}, \mathcal{V}^{t+3}) \leq Cr^{dt} \quad (1.9)$$

Here $r < 1$ and d are derived explicitly and depend only on the parameters associated with the problem, while C is an easily computable constant which depends on initial conditions as well as the parameters. In particular, if we let $\mathcal{U}^0 = (1, 1, 1, 1)$ and $\mathcal{V}^0 \sim \pi$, then by (1.9) we have a bound on the distance to equilibrium. We obtain a result similar to (1.9) for the case $n > 4$ (with different values for r , C and d), where we use a slightly modified version of the previous sampler. Our result for the $n = 4$ case holds true for a wide selection of parameter values, while the $n > 4$ case requires an additional precondition which holds true whenever the parameters $\{a_i\}$ satisfy some constraints (which are difficult to state in brevity, but are fulfilled in the special case when all a_i are equal and greater than some fixed number).

One more problem considered in this thesis is based on [3], where A.L. Gibbs proposes a stochastic image restoration model which works on the assumption that random noise (with a truncated normal distribution) has been superimposed on a greyscale image made up of N pixels. It is also assumed that in the pre-noise image, the darkness of a pixel is related to that of its nearest n neighbouring pixels. The

author defines a Markov chain X_t that converges to the posterior distribution π of the image, and provides a bound of the form

$$T_W \leq CN \log \left(\frac{N}{\epsilon} \right) \tag{1.10}$$

for the Wasserstein mixing time T_W , which is defined as the minimum time such that $d_W(X_t, \pi) < \epsilon$ for all $t > T$. In Section 5.2 we obtain an analogous result for convergence in total variation. In other words, we make use of (1.10) to show that if τ is the minimum time such that $d_{TV}(X_t, \pi) < \epsilon$ for all $t \geq \tau$, then τ is also $O(N \log(N))$. As was already stated (and also described in [4]), there is a fundamental difference between the metric d_{TV} and d_W . Depending on the purpose or application, one may prove more useful than other.

2 Hierarchical Gibbs Sampler

In this part we study methods that may be used by applied statisticians to assess the validity of this type of hierarchical models (see introduction for examples of applications in numerous areas). The problems considered in Section 2.1 and Section 4 can be motivated by the following Bayesian-inference scenario: suppose we are given a known quantity $x \in \mathbb{R}^+$ along with the information that x was sampled from a $\Gamma(a_1, u_1)$ probability distribution, defined by the probability density function

$$f(z) = \frac{u_1^{a_1}}{\Gamma(a_1)} z^{a_1-1} e^{-zu_1}$$

where $a_1 > 0$ is given. The inverse scale parameter u_1 is itself the product of random sampling from an independent $\Gamma(a_2, u_2)$ distribution. Once again it is assumed 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 of (x, u_1, \dots, u_n) is therefore proportional to

$$p(x, z_1, \dots, z_n) \propto x^{a_1-1} \left(\prod_{i=1}^n z_i^{a_i+a_{i+1}-1} \right) \exp \left(\sum_{i=1}^{n+1} -z_i z_{i-1} \right) \quad (2.1)$$

where for convenience we define $z_0 := x$ and $z_{n+1} := b$. We conclude from (2.1) that for $1 \leq i \leq n$, the conditional densities will be

$$p\left(z_i \mid x, z_{j \neq i}\right) \propto z_i^{a_i + a_{i+1} - 1} \exp\left(-z_i(z_{i-1} + z_{i+1})\right) \quad (2.2)$$

By (2.2), the distribution of u_i given everything else is

$$u_i \mid x, u_{j \neq i} \sim \Gamma\left(a_i + a_{i+1}, u_{i-1} + u_{i+1}\right)$$

The resulting posterior distribution of (u_1, \dots, u_n) (i.e. given x as well as all other parameters) is therefore defined by the density function

$$g(z_1, \dots, z_n) \propto \left(\prod_{i=1}^n z_i^{a_i + a_{i+1} - 1}\right) \exp\left(\sum_{i=1}^{n+1} -z_i z_{i-1}\right) \quad (2.3)$$

Should one wish to sample from it, however, it would be quite challenging (in particular for large values of n) to do so directly due to its complicated structure. More importantly, our work allows us to better understand the mixing properties of the hierarchical sampler.

2.1 Random dynamical systems

Our first attempt at the problem stated above will be for the case $n = 3$. We start with the relations $x \sim \Gamma(a_1, u)$, $u \sim \Gamma(a_2, v)$, $v \sim \Gamma(a_3, w)$ and $w \sim \Gamma(a_4, b)$. The joint density function (2.3) becomes

$$f(x, u, v, w) \propto u^{a_1} x^{a_1 - 1} e^{-xu} v^{a_2} u^{a_2 - 1} e^{-uv} w^{a_3} v^{a_3 - 1} e^{-vw} b^{a_4} w^{a_4 - 1} e^{-wb} \quad (2.4)$$

and the conditional densities become

$$f(u | x, v, w) = \frac{f(x, u, v, w)}{\int f(x, u, v, w) du} \propto u^{a_1+a_2-1} e^{-u(x+v)} \quad (2.5)$$

with similar equations for v and w . Recall that this implies $u | x, v, w \sim \Gamma(a_1 + a_2, x + v)$,

and similarly $v | x, u, w \sim \Gamma(a_2 + a_3, u + w)$ and $w | x, u, v \sim \Gamma(a_3 + a_4, v + b)$.

The Markov chain which sequentially updates u, v and w , in that particular order,

is defined by

$$(u_{t+1}, v_{t+1}, w_{t+1}) := \left(\frac{\gamma_1^{t+1}}{x + v_t}, \frac{\gamma_2^{t+1}}{\frac{\gamma_1^{t+1}}{x+v_t} + w_t}, \frac{\gamma_3^{t+1}}{\frac{\gamma_2^{t+1}}{\frac{\gamma_1^{t+1}}{x+v_t} + w_t} + b} \right) \quad (2.6)$$

where $\gamma_1^{t+1} \sim \Gamma(a_1 + a_2, 1)$, $\gamma_2^{t+1} \sim \Gamma(a_2 + a_3, 1)$ and $\gamma_3^{t+1} \sim \Gamma(a_3 + a_4, 1)$ are independent. Equation (2.6) suggests that it would suffice to consider the 2 dimensional

process

$$(v_{t+1}, w_{t+1}) := \left(\frac{\gamma_2^{t+1}}{\frac{\gamma_1^{t+1}}{x+v_t} + w_t}, \frac{\gamma_3^{t+1}}{\frac{\gamma_2^{t+1}}{\frac{\gamma_1^{t+1}}{x+v_t} + w_t} + b} \right)$$

which is also a Markov chain. In fact, for $t \geq 1$ we observe that

$$v_{t+1} = \frac{\gamma_2^{t+1}}{\frac{\gamma_1^{t+1}}{x+v_t} + w_t} = \frac{\gamma_2^{t+1}}{\frac{\gamma_1^{t+1}}{x+v_t} + \frac{\gamma_3^t}{b+v_t}}$$

It will follow from Corollary 2.4 that an analysis of the behaviour of (2.6) can be done by considering the Markov chain defined by the iterated function system

$\{f_t\}_{t=1}^\infty$ where

$$f_t(v) := \frac{\gamma_2^t}{\frac{\gamma_1^t}{x+v} + \frac{\gamma_3^t}{b+v}} \quad (2.7)$$

Consistent with the notation in [5], we let $F_t(v) := f_1 \circ f_2 \circ \dots \circ f_t(v)$ and $F_\infty(v) := \lim_{t \rightarrow \infty} F_t(v)$, if the limit exists. Conditions for the existence of this limit are discussed in the aforementioned article, as well as conditions that imply $F_\infty(v)$ is independent of v (in which case it will be denoted by F_∞). A sufficient condition for this is that the random iterated function system is 1-contractive, as per Definition 1.1. This is given in Theorem 1 of [5] (with $\Phi(.,.)$ identically 1, and $x = y$), which we will re-phrase in a form more suitable to our problem.

Theorem 2.1. *[5] Suppose that $\{f_t\}$ is 1-contractive. Then*

$$\mathbb{E} \left[|F_n(x) - F_\infty| \right] \leq \frac{r^n}{1-r} \mathbb{E} \left[|f(x) - x| \right]$$

where r is a bound as given in (1.6).

It will be established in the proof of Theorem 2.2 that this is the case for a 'multi-step' version of the system $\{F_t(v)\}$. Hence it will follow that $\{F_t(v)\}$ itself must also converge to some F_∞ .

We can now state our main result for this iterated function system.

Theorem 2.2. *Suppose that $a_1 + a_4 > 2$. Then for $1 \leq l < k$ and for all $t \geq 0$*

$$d_W(F_{kt+l}(v), F_\infty) \leq \frac{r^t}{1-r} \left(\frac{|x-b|}{1-\alpha} + v\alpha^l \right)$$

with $\alpha = \frac{a_2+a_3}{a_1+a_2+a_3+a_4-1}$, $r = \frac{1}{2}(1+\alpha)$, $k = \left\lceil \frac{1}{2} \left(\frac{64\alpha x^3}{b^3(1+\alpha)^2(1-\alpha)^7(1-\rho)^2} + 3 \right) \right\rceil$ and $\rho = \max \left(\alpha, \mathbb{E} \left[\left(\frac{\gamma_2}{\gamma_1+\gamma_3} \right)^2 \right] \right)$.

Note from the following lemma that $\rho < 1$. A similar result for the Markov chain in (2.6) is given in Corollary 2.4.

Lemma 2.3. *Let $\gamma \sim \Gamma(a, b)$. Then if $a > 1$, $\mathbb{E} \left[\frac{1}{\gamma} \right] = \frac{b}{a-1}$ and if $a > 2$, $\mathbb{E} \left[\frac{1}{\gamma^2} \right] = \frac{b^2}{(a-1)(a-2)}$.*

The proof of Lemma 2.3 can be found in [26].

Proof of Theorem 2.2. Let us assume w.l.o.g. that $x \geq b$ (if $b > x$, switching these two constants in this proof starting with the first inequality below, it can be easily seen that everything follows as before), so that

$$\begin{aligned} f'_t(v) &= \frac{\gamma_2^t}{\left(\frac{\gamma_1^t}{x+v} + \frac{\gamma_3^t}{b+v} \right)^2} \cdot \left(\frac{\gamma_1^t}{(x+v)^2} + \frac{\gamma_3^t}{(b+v)^2} \right) \\ &\leq \frac{\gamma_2^t}{\left(\frac{\gamma_1^t}{x+v} + \frac{\gamma_3^t}{b+v} \right)^2} \cdot \left(\frac{\gamma_1^t}{x+v} + \frac{\gamma_3^t}{b+v} \right) \cdot \frac{1}{b+v} \\ &= f_t(v) \cdot \frac{1}{b+v} \end{aligned} \tag{2.8}$$

Applying the chain rule, we get

$$\begin{aligned} (f_1 \circ f_2 \circ \dots \circ f_k(v))' &= f'_1(f_2 \circ f_3 \circ \dots \circ f_k(v)) \cdot f'_2(f_3 \circ \dots \circ f_k(v)) \cdots f'_k(v) \\ &= f'_1(v_{2,k}) \cdot f'_2(v_{3,k}) \cdots f'_k(v) \\ &\leq \frac{f_1(v_{2,k})}{b+v} \cdot \frac{v_{2,k}}{b+v_{2,k}} \cdots \frac{v_{k,k}}{b+v_{k,k}} \end{aligned}$$

where $v_{i,k} = f_i \circ f_{i+1} \circ \dots \circ f_k(v)$. Suppose first that $v \geq \frac{2x\alpha}{(1-\alpha)^2}$, where $\alpha =$

$\frac{a_2+a_3}{a_1+a_2+a_3+a_4-1}$, and let

$$h_t(v) := \frac{\gamma_2^t}{\gamma_1^t + \gamma_3^t} \cdot (x+v)$$

and

$$H_k(v) := h_1 \circ \dots \circ h_k(v)$$

Observe that $f_t \leq h_t$ for all t . Since both are monotone increasing (for all t), it follows that $F_t \leq H_t$ for all t . Then for any $k \geq 1$,

$$\begin{aligned} \mathbb{E} \left[(f_1 \circ f_2 \circ \dots \circ f_k(v))' \right] &\leq \mathbb{E} \left[\frac{f_1(v_{2,k})}{b+v} \right] \\ &\leq \mathbb{E} \left[\frac{H_k(v)}{b+v} \right] \\ &= \frac{1}{b+v} \mathbb{E} \left[v \cdot \left(\prod_{j=1}^k \frac{\gamma_2^j}{\gamma_1^j + \gamma_3^j} \right) + x \cdot \sum_{i=1}^k \prod_{j=1}^i \frac{\gamma_2^j}{\gamma_1^j + \gamma_3^j} \right] \\ &= \frac{1}{b+v} \left(v\alpha^k + x\alpha \frac{1-\alpha^k}{1-\alpha} \right) \\ &\leq \alpha + \frac{x\alpha}{(1-\alpha)(b+v)} \\ &\leq \alpha + \frac{x\alpha}{(1-\alpha)v} \\ &\leq \frac{1+\alpha}{2} \end{aligned} \tag{2.9}$$

Hence for $v \geq \frac{2x\alpha}{(1-\alpha)^2}$, (2.9) tells us that $f_1 \circ f_2 \circ \dots \circ f_k(v)$ 1-contractive, for any value of k . We will now show that if $v < \frac{2x\alpha}{(1-\alpha)^2}$ and k is sufficiently large, then

$f_1 \circ f_2 \circ \dots \circ f_k(v)$ is once again 1-contractive. Observe that

$$\begin{aligned}
\mathbb{E} \left[(f_1 \circ f_2 \circ \dots \circ f_k(v))' \right] &\leq \mathbb{E} \left[\frac{f_1(v_{2,k})}{b+v} \cdot \frac{v_{2,k}}{b+v_{2,k}} \dots \frac{v_{k,k}}{b+v_{k,k}} \right] \\
&\leq \mathbb{E} \left[\frac{h_1(\hat{v}_{2,k})}{b+v} \cdot \frac{\hat{v}_{2,k}}{b+\hat{v}_{2,k}} \dots \frac{\hat{v}_{k,k}}{b+\hat{v}_{k,k}} \right] \\
&= \frac{1}{b+v} \left\| H_k(v) \cdot \frac{\hat{v}_{2,k}}{b+\hat{v}_{2,k}} \dots \frac{\hat{v}_{k,k}}{b+\hat{v}_{k,k}} \right\|_1
\end{aligned} \tag{2.10}$$

where $\hat{v}_{i,k} = h_i \circ \dots \circ h_k(v)$, and the norm used in the last line is the L_1 -norm.

Observe that all factors inside the norm, except possibly for $H_k(v)$, are less than

1. Our aim is to show that for k sufficiently large, these values will accumulate to

make (2.10) less than one. Note that

$$\begin{aligned}
H_k(v) &= v \cdot \left(\prod_{j=1}^k \frac{\gamma_2^j}{\gamma_1^j + \gamma_3^j} \right) + x \cdot \sum_{i=1}^k \prod_{j=1}^i \frac{\gamma_2^j}{\gamma_1^j + \gamma_3^j} \\
&\leq \frac{2x\alpha}{(1-\alpha)^2} \cdot \left(\prod_{j=1}^k \frac{\gamma_2^j}{\gamma_1^j + \gamma_3^j} \right) + Z_\infty \\
&\leq Z_\infty \left(\frac{2\alpha}{(1-\alpha)^2} + 1 \right)
\end{aligned}$$

where $Z_\infty = x \cdot \sum_{i=1}^\infty \prod_{j=1}^i \frac{\gamma_2^j}{\gamma_1^j + \gamma_3^j}$. Similarly, for all $1 \leq j \leq k$

$$\frac{\hat{v}_{j,k}}{b + \hat{v}_{j,k}} \leq \frac{Z_{\infty,j} \left(\frac{2\alpha}{(1-\alpha)^2} + 1 \right)}{b + Z_{\infty,j} \left(\frac{2\alpha}{(1-\alpha)^2} + 1 \right)}$$

where $Z_{\infty,j} = x \cdot \sum_{i=j}^\infty \prod_{l=j}^i \frac{\gamma_2^l}{\gamma_1^l + \gamma_3^l}$ has the same distribution as Z_∞ . Then by applying

Holder's inequality twice, we get

$$\begin{aligned}
& \mathbb{E} \left[(f_1 \circ f_2 \circ \dots \circ f_k(v))' \right] \\
& \leq \frac{\left(\frac{2\alpha}{(1-\alpha)^2} + 1 \right)}{b+v} \left\| Z_\infty \frac{\hat{v}_{2,k}}{b + \hat{v}_{2,k}} \dots \frac{\hat{v}_{k,k}}{b + \hat{v}_{k,k}} \right\|_1 \\
& \leq \frac{\left(\frac{2\alpha}{(1-\alpha)^2} + 1 \right)}{b} \left\| Z_\infty \frac{Z_{\infty,2} \left(\frac{2\alpha}{(1-\alpha)^2} + 1 \right)}{b + Z_{\infty,2} \left(\frac{2\alpha}{(1-\alpha)^2} + 1 \right)} \dots \frac{Z_{\infty,k} \left(\frac{2\alpha}{(1-\alpha)^2} + 1 \right)}{b + Z_{\infty,k} \left(\frac{2\alpha}{(1-\alpha)^2} + 1 \right)} \right\|_1 \\
& \leq \frac{\left(\frac{2\alpha}{(1-\alpha)^2} + 1 \right)}{b} \|Z_\infty\|_2 \left\| \left(\frac{Z_\infty \left(\frac{2\alpha}{(1-\alpha)^2} + 1 \right)}{b + Z_\infty \left(\frac{2\alpha}{(1-\alpha)^2} + 1 \right)} \right)^2 \right\|_{k-1}^{\frac{k-1}{2}} \\
& \leq \frac{2}{b(1-\alpha)^2} \|Z_\infty\|_2 \sqrt{\mathbb{E} \left[\left(\frac{Z_\infty \left(\frac{2\alpha}{(1-\alpha)^2} + 1 \right)}{b + Z_\infty \left(\frac{2\alpha}{(1-\alpha)^2} + 1 \right)} \right)^{2(k-1)} \right]}
\end{aligned} \tag{2.11}$$

In order to find a bound for the term on the right-hand side of the last line of (2.11),

we let $Y = Z_\infty \left(\frac{2\alpha}{(1-\alpha)^2} + 1 \right)$ and observe that

$$\mathbb{E} \left[\left(\frac{Y}{b+Y} \right)^{2(k-1)} \right] = \int_0^1 \mathbb{P} \left[\left(\frac{Y}{b+Y} \right)^{2(k-1)} > y \right] dy$$

Note that also

$$\begin{aligned}
& \left(\frac{Y}{b+Y} \right)^{2(k-1)} > y \\
& \Leftrightarrow \left(\frac{Y}{b+Y} \right) > y^{\frac{1}{2(k-1)}} \\
& \Leftrightarrow Y \left(1 - y^{\frac{1}{2(k-1)}} \right) > by^{\frac{1}{2(k-1)}} \\
& \Leftrightarrow Y > \frac{by^{\frac{1}{2(k-1)}}}{\left(1 - y^{\frac{1}{2(k-1)}} \right)}
\end{aligned}$$

Hence

$$\begin{aligned}
\mathbb{E} \left[\left(\frac{Y}{b+Y} \right)^{2(k-1)} \right] &= \int_0^1 \mathbb{P} \left[Y > \frac{by^{\frac{1}{2(k-1)}}}{(1-y^{\frac{1}{2(k-1)}})} \right] dy \\
&\leq \int_0^1 \mathbb{E}[Y] \frac{(1-y^{\frac{1}{2(k-1)}})}{by^{\frac{1}{2(k-1)}}} dy \\
&\leq \left(\frac{2\alpha}{(1-\alpha)^2} + 1 \right) \frac{\alpha x}{1-\alpha} \int_0^1 \frac{1}{b} y^{\frac{-1}{2(k-1)}} - \frac{1}{b} dy \\
&= \left(\frac{2\alpha}{(1-\alpha)^2} + 1 \right) \frac{\alpha x}{1-\alpha} \frac{1}{b} \left(\frac{1}{1-\frac{1}{2(k-1)}} - 1 \right) \\
&\leq \frac{2\alpha x}{b(1-\alpha)^3} \frac{1}{2k-3}
\end{aligned}$$

The random variable $\frac{\gamma_2}{\gamma_1+\gamma_3}$ is a ratio of independent gamma random variables, and therefore by Lemma 2.3

$$\mathbb{E} \left[\left(\frac{\gamma_2}{\gamma_1 + \gamma_3} \right)^2 \right] = \frac{(a_2 + a_3)(a_2 + a_3 + 1)}{(a_1 + a_2 + a_3 + a_4 - 1)(a_1 + a_2 + a_3 + a_4 - 2)}$$

So in particular since $a_1 + a_4 > 2$, the above expectation is less than 1. Let

$$Z_k = x \cdot \sum_{i=1}^k \prod_{j=1}^i \frac{\gamma_2^j}{\gamma_1^j + \gamma_3^j} = x \cdot \sum_{i=1}^k \prod_{j=1}^i \beta_j$$

with $\beta_j := \frac{\gamma_2^j}{\gamma_1^j + \gamma_3^j}$, and $\rho = \max \left(\alpha, \mathbb{E} \left[\left(\frac{\gamma_2}{\gamma_1 + \gamma_3} \right)^2 \right] \right)$. Then

$$\begin{aligned}
\mathbb{E} [Z_k^2] &\leq 2x^2 \left(\frac{\rho}{1-\rho} + \frac{\rho^2}{1-\rho} + \dots + \frac{\rho^k}{1-\rho} \right) \\
&\leq 2x^2 \left(\frac{1}{1-\rho} \right)^2
\end{aligned}$$

Hence by the MCT

$$\|Z_\infty\|_2 \leq \sqrt{2}x \left(\frac{1}{1-\rho} \right)$$

Therefore, if we let

$$k = \left\lceil \frac{1}{2} \left(\frac{64\alpha x^3}{b^3(1+\alpha)^2(1-\alpha)^7(1-\rho)^2} + 3 \right) \right\rceil$$

we once again get $\mathbb{E} \left[(f_1 \circ f_2 \circ \dots \circ f_k(v))' \right] \leq \frac{1+\alpha}{2}$.

Now set $r = \frac{1+\alpha}{2}$, then by Theorem 2.1, for any $0 \leq l < k$ we have

$$\begin{aligned} d_W(F_{kt+l}(v), F_\infty) &\leq \frac{r^t}{1-r} \mathbb{E} \left[|f_1 \circ \dots \circ f_{k+l}(v) - f_1 \circ \dots \circ f_l(v)| \right] \\ &\leq \frac{r^t}{1-r} \left(\frac{|x-b|}{1-\alpha} + v\alpha^l \right) \end{aligned}$$

where we have used the fact that

$$b \cdot \sum_{i=1}^m \prod_{j=1}^i \beta_j + v \prod_{j=1}^m \beta_j \leq f_1 \circ \dots \circ f_m(v) \leq x \cdot \sum_{i=1}^m \prod_{j=1}^i \beta_j + v \prod_{j=1}^m \beta_j$$

□

Now if we consider the taxicab metric on \mathbb{R}^3 and let $d_{\bar{W}}$ be the corresponding Wasserstein metric, then we can relate the result of Theorem 2.2 to (2.6) through the following corollary

Corollary 2.4. *Under the same conditions as in Theorem 2.2, if $Y^t = (Y_1^t, Y_2^t, Y_3^t)$ is a copy of (2.6),*

$$d_{\bar{W}}(Y^{kt+l+1}, \pi) \leq \frac{r^t}{1-r} \left(\frac{|x-b|}{1-\alpha} + Y_2^0 \alpha^l \right) \left(\frac{a_1 + a_2}{x^2} + \alpha \frac{\max\{x, b\}}{\min\{x, b\}} \left(\frac{a_3 + a_4}{b^2} + 1 \right) \right)$$

where π is the equilibrium distribution of this Markov chain.

Proof. Note that $F_t(Y_2^0) \sim Y_2^t$ and $F_\infty \sim \pi_2$ (the marginal distribution of π in the second variable), hence we can construct a pair of random variables $\omega \sim \pi_2$ and $\phi \sim Y_2^{kt+l}$ such that $\mathbb{E}[|\omega - \phi|] \leq \frac{r^t}{1-r} \left(\frac{|x-b|}{1-\alpha} + Y_2^0 \alpha^l \right)$. Let $\gamma_1 \sim \Gamma(a_1 + a_2, 1)$, $\gamma_2 \sim \Gamma(a_2 + a_3, 1)$, $\gamma_3 \sim \Gamma(a_3 + a_4, 1)$ and $\gamma'_3 \sim \Gamma(a_3 + a_4, 1)$ be independent of each other, as well as independent of $\{Y_2^{kt+l}\}$, ϕ and ω . Then

$$\left(Y_1^{kt+l+1}, Y_2^{kt+l+1}, Y_3^{kt+l+1} \right) \sim \left(\frac{\gamma_1}{\phi + x}, f^*(\phi), \frac{\gamma'_3}{f^*(\phi) + b} \right)$$

where $f^*(y) = \gamma_2 / (\gamma_1 / (y + x) + \gamma_3 / (y + b))$ for $y \in \mathbb{R}_+$. Note also that

$$(\pi_1, \pi_2, \pi_3) \sim \left(\frac{\gamma_1}{\pi_2 + x}, f^*(\pi_2), \frac{\gamma'_3}{f^*(\pi_2) + b} \right)$$

It immediately follows that $\mathbb{E} \left[\left| \frac{\gamma_1^{kt+l+1}}{\phi+x} - \frac{\gamma_1^{kt+l+1}}{\omega+x} \right| \right] \leq \frac{r^t}{1-r} \left(\frac{|x-b|}{1-\alpha} + Y_2^0 \alpha^l \right) \left(\frac{a_1+a_2}{x^2} \right)$.

Now observe from (2.8) and the mean value theorem that for some $\min\{\omega, \phi\} \leq z \leq \max\{\omega, \phi\}$,

$$\begin{aligned} \mathbb{E} \left[|f^*(\omega) - f^*(\phi)| \right] &\leq \mathbb{E} \left[\frac{f^*(z)}{\min\{x, b\} + z} |\omega - \phi| \right] \\ &\leq \mathbb{E} \left[\frac{\gamma_2}{\gamma_1 + \gamma_3} \frac{\max\{x, b\} + z}{\min\{x, b\} + z} |\omega - \phi| \right] \\ &\leq \frac{r^t}{1-r} \left(\frac{|x-b|}{1-\alpha} + Y_2^0 \alpha^l \right) \alpha \frac{\max\{x, b\}}{\min\{x, b\}} \end{aligned}$$

Lastly, observe that

$$\begin{aligned} \mathbb{E} \left[\left| \frac{\gamma'_3}{f^*(\phi) + b} - \frac{\gamma'_3}{f^*(\omega) + b} \right| \right] &\leq \frac{(a_3 + a_4)}{b^2} \mathbb{E} \left[|f^*(\omega) - f^*(\phi)| \right] \\ &\leq \frac{(a_3 + a_4)}{b^2} \frac{r^t}{1-r} \left(\frac{|x-b|}{1-\alpha} + Y_2^0 \alpha^l \right) \alpha \frac{\max\{x, b\}}{\min\{x, b\}} \end{aligned}$$

This proves the statement of the corollary. \square

The condition $a_1 + a_4 > 2$ in the statement of Theorem 2.2 is partly there because it simplifies the proof, but to some extent it is also a necessity for convergence in d_W to occur, as can be seen from the following corollary.

Corollary 2.5. *Suppose $a_1 + a_4 < 1$ and $a_1 + a_2 + a_3 + a_4 > 1$. Then if $\{X_t\}$ is defined by (2.7) and X_∞ is in equilibrium*

$$d_W(X_t, X_\infty) = \infty$$

Proof. From equation (1.7) we can infer that the equilibrium distribution of $\{X_t\}$ has a density function

$$G(v) \propto \frac{v^{a_2 + a_3 - 1}}{(x + v)^{a_1 + a_2} (v + b)^{a_3 + a_4}} \quad (2.12)$$

Hence if X_∞ is distributed according to (2.12), $\mathbb{E}[X_\infty] = \infty$. From the inequality

$$\begin{aligned} \mathbb{E}[X_t] &\leq \mathbb{E}\left[\frac{\gamma_2^t}{\gamma_1^t + \gamma_3^t} (X_{t-1} + \max(x, b))\right] \\ &= \frac{a_2 + a_3}{a_1 + a_2 + a_3 + a_4 - 1} (\mathbb{E}[X_{t-1}] + \max(x, b)) \end{aligned}$$

it is clear that $\mathbb{E}[X_t] < \infty$ for all t . It follows that

$$\mathbb{E}[|X_t - X_\infty|] = \infty$$

□

The result of Theorem 2.2 gives us a bound on the rate of convergence of our Gibbs sampler for the case $n = 3$. The proof is adapted to this problem and uses

some ad hoc methods that don't translate easily to other problems. In particular, bounding the absolute value of a 'derivative' in higher dimensions becomes problematic (assuming we can settle for a convenient metric which will induce a derivative) which is why we pursue a different approach in Section 4.

As a practical example, let us consider the case when $x = 2$, $b = 1$ and $a_i = i + 1$. Then a simple computation shows that $k = 3294$, and therefore based on this bound a Markov chain starting at $v = 1$ would need at most 180,000 iterations to be within 10^{-5} of equilibrium in the Wasserstein metric.

2.2 Multiple data points

An extension to the problem of Section 2.1 is the Bayesian estimator problem where the data set $\vec{x} := x_1, \dots, x_J$ consists of J sample points. These are sampled so that apriori $x_i \sim \Gamma(a_i, b_i)$, and co-ordinates are drawn independently. As in the previous problem, the shape parameters a_i are given, and the scale parameters b_i are themselves sampled from independent $\Gamma(\bar{a}_i, c)$ distributions. In addition to this, we will also assume that \bar{a}_i are given while $c \sim \Gamma(a_{J+1}, d)$. The joint posterior density function satisfies the relation

$$\begin{aligned}
 & h(x_1, \dots, x_J, b_1, \dots, b_J, c) && (2.13) \\
 & \propto \left(\prod_{i=1}^J b_i^{a_i} x_i^{a_i-1} \exp(-b_i x_i) \right) \left(\prod_{i=1}^J c^{\bar{a}_i} b_i^{\bar{a}_i-1} \exp(-b_i c) \right) c^{a_{J+1}-1} \exp(-cd)
 \end{aligned}$$

and therefore we arrive at the following marginal density functions

$$\begin{aligned} h_i \left(b_i \mid \vec{x}, b_{k \neq i}, c \right) &\propto b_i^{a_i + \bar{a}_i - 1} \exp(-b_i(x_i + c)) \\ h_c \left(c \mid \vec{x}, b \right) &\propto c^{a_{J+1} - 1 + \sum \bar{a}_i} \exp\left(-c \left(d + \sum b_i\right)\right) \end{aligned}$$

From this we infer that $b_i \mid \vec{x}, b_{k \neq i}, c \sim \Gamma(a_i + \bar{a}_i, x_i + c)$ and

$$c \mid \vec{x}, b \sim \Gamma(a_{J+1} + \sum \bar{a}_i, d + \sum b_i).$$

Let π be the probability density associated with the conditional density function $h_{bc}(b_1, \dots, b_J, c \mid x_1, \dots, x_J)$. We consider the iterated function system given by

$$f_{t+1} \left(b_1^t, \dots, b_J^t, c^t \right) := \left(\frac{\gamma_1^{t+1}}{x_1 + c^t}, \dots, \frac{\gamma_J^{t+1}}{x_J + c^t}, \frac{\gamma_{J+1}^{t+1}}{d + \sum b_i^{t+1}} \right) \quad (2.14)$$

which has as its unique invariant distribution the probability measure π (in other words, if $Y \sim \pi$, then $f_t(Y) \sim \pi$). Here $\gamma_i^{t+1} \sim \Gamma(a_i + \bar{a}_i, 1)$ are independent over time and co-ordinate wise for all $i = 1, \dots, J$, and similar independence is extended to $\gamma_{J+1} \sim \Gamma(a_{J+1} + \sum \bar{a}_i, 1)$. It can be seen immediately that

$$c^{t+1} = \frac{\gamma_{J+1}^{t+1}}{d + \frac{\gamma_1^{t+1}}{x_1 + c^t} + \dots + \frac{\gamma_J^{t+1}}{x_J + c^t}}$$

This motivates the definition of the simpler, one-variable random dynamical system

$\{g_t\}_{t=0}^\infty$ defined by

$$g_{t+1}(c) := \frac{\gamma_{J+1}^{t+1}}{d + \frac{\gamma_1^{t+1}}{x_1 + c} + \dots + \frac{\gamma_J^{t+1}}{x_J + c}} \quad (2.15)$$

Similar to the notation in the previous section, we define the backwards iterated function system $G_t(v) := g_1 \circ \dots \circ g_t(v)$ and $G_\infty(v) := \lim_{t \rightarrow \infty} G_t(v)$. Our result

for this system is given in the following theorem. The proof is very similar to the one of Theorem 2.2.

Theorem 2.6. *For the iterated function system defined above, and for all $t \geq 0$*

$$\mathbb{E} \left[|G_{kt+j+1}(c) - G_\infty| \right] \leq \left(\frac{2\mu_{J+1}}{d} \right) \left(\frac{1}{2} \right)^{t-1}$$

where $k = \max \left\{ 1, \left\lceil 1 + \frac{\ln \left(\frac{d \cdot \min \{x_i\}}{2\mu_{J+1}} \right)}{\ln \left(\frac{\mu_{J+1}}{d \cdot \min \{x_i\} + \mu_{J+1}} \right)} \right\rceil \right\}$, $1 \leq j \leq k - 1$ and $\mu_{J+1} = \mathbb{E} [\gamma_{J+1}]$.

Proof. We are interested in bounding the derivative of this function, as was done in Section 2.1. Observe that

$$\begin{aligned} g'_{t+1}(c) &= \frac{\gamma_{J+1}^{t+1}}{\left(d + \frac{\gamma_1^{t+1}}{x_1+c} + \dots + \frac{\gamma_J^{t+1}}{x_J+c} \right)^2} \left(\frac{\gamma_1^{t+1}}{(x_1+c)^2} + \dots + \frac{\gamma_J^{t+1}}{(x_J+c)^2} \right) \\ &\leq \frac{\gamma_{J+1}^{t+1}}{\left(d + \frac{\gamma_1^{t+1}}{x_1+c} + \dots + \frac{\gamma_J^{t+1}}{x_J+c} \right)^2} \left(\frac{\gamma_1^{t+1}}{(x_1+c)} + \dots + \frac{\gamma_J^{t+1}}{(x_J+c)} \right) \frac{1}{\min \{x_i\} + c} \\ &\leq g_{t+1}(c) \frac{1}{\min \{x_i\} + c} \end{aligned} \tag{2.16}$$

Now let $k \geq 1$ be some integer, and define $\bar{g}_0(c) := c$ and $\bar{g}_t(c) := g_{k(t-1)+1} \circ g_{k(t-1)+2} \circ \dots \circ g_{kt}(c)$ for $t \geq 1$. Then using the notation $c_{k(t-1)+i,kt} := g_{k(t-1)+i} \circ$

$\dots \circ g_{kt}(c)$ for $1 \leq i \leq k$ we conclude by the chain rule and (2.16) that

$$\begin{aligned}
& \bar{g}'_t(c) \\
& \leq \frac{g_{k(t-1)+1}(c_{k(t-1)+2,kt})}{\min\{x_i\} + c_{k(t-1)+2,kt}} \frac{g_{k(t-1)+2}(c_{k(t-1)+3,kt})}{\min\{x_i\} + c_{k(t-1)+3,kt}} \dots \frac{g_{kt}(c)}{\min\{x_i\} + c} \\
& = \frac{\bar{g}_t(c)}{\min\{x_i\} + c} \cdot \frac{c_{k(t-1)+2,kt}}{\min\{x_i\} + c_{k(t-1)+2,kt}} \cdot \frac{c_{k(t-1)+3,kt}}{\min\{x_i\} + c_{k(t-1)+3,kt}} \dots \frac{c_{kt,kt}}{\min\{x_i\} + c_{kt,kt}} \\
& \leq \frac{\gamma_{J+1}^{k(t-1)+1}}{d(\min\{x_i\} + c)} \cdot \frac{\gamma_{J+1}^{k(t-1)+2}}{d} \dots \frac{\gamma_{J+1}^{kt}}{d} \\
& \quad \min\{x_i\} + \frac{\gamma_{J+1}^{k(t-1)+2}}{d} \quad \min\{x_i\} + \frac{\gamma_{J+1}^{kt}}{d}
\end{aligned}$$

Then by Jensen's inequality $\mathbb{E}[\bar{g}'_t(c)] \leq \frac{\mu_{J+1}}{d(\min\{x_i\} + c)} \left(\frac{\mu_{J+1}}{d \cdot \min\{x_i\} + \mu_{J+1}} \right)^{k-1}$ where

$\mu_{J+1} = \mathbb{E}[\gamma_{J+1}] = a_{J+1} + \sum \bar{a}_i$. Therefore, taking $k = \left\lceil 1 + \frac{\ln\left(\frac{d \cdot \min\{x_i\}}{2\mu_{J+1}}\right)}{\ln\left(\frac{\mu_{J+1}}{d \cdot \min\{x_i\} + \mu_{J+1}}\right)} \right\rceil$ if $d \cdot \min\{x_i\} \leq 2\mu_{J+1}$ and $k = 1$ otherwise, we conclude by Theorem 1 of [5] that

for $1 \leq j \leq k - 1$

$$\begin{aligned}
& \mathbb{E} \left[\left| G_{kt+j+1}(c) - G_\infty \right| \right] \\
& = \mathbb{E} \left[\left| G_{kt} \left(g_{kt+1} \circ \dots \circ g_{kt+j} \left(g_{kt+j+1}(c) \right) \right) - G_\infty \right| \right] \\
& \leq \left(\mathbb{E} \left[\left| g_{kt+1} \circ \dots \circ g_{kt+j} \left(g_{kt+j+1}(c) \right) - g_{kt+j+1}(c) \right| \right] \right) \left(\frac{1}{2} \right)^{t-1} \\
& \leq \left(\frac{2\mu_{J+1}}{d} \right) \left(\frac{1}{2} \right)^{t-1}
\end{aligned}$$

□

A notable difference from Theorem 2.2, is that the bound in Theorem 2.6 is

uniform and hence independent of the initial state. This is due to the fact that the random function in (2.15) is bounded in expectation.

The problem in Theorem 2.6 has a hierarchical 2-level depth, and a natural extension would be to consider a similar 3-level hierarchical sampler where apriori $d \sim \Gamma(a_{J+2}, z)$ for some $a_{J+2}, z \in \mathbb{R}^+$ that are known. With derivations similar to those given above (and also in Section 2.1), and assuming a similar notation, we can conclude that the iterated function system defined by

$$f_{t+1}(c) = \frac{\gamma_{J+1}^{t+1}}{\frac{\gamma_{J+2}^{t+1}}{z+c} + \frac{\gamma_1^{t+1}}{x_1+c} + \dots + \frac{\gamma_J^{t+1}}{x_J+c}} \quad (2.17)$$

will converge to the relevant equilibrium. Then from the inequality

$$f'_{t+1}(c) \leq \frac{f_{t+1}(c)}{\min\{x_1, \dots, x_J, z\} + c}$$

it is easy to see that we can replicate the proof of Theorem 2.2 to obtain a similar result.

A further possible extension to this problem would be to consider the same Bayesian estimator defined at the beginning of this section, with an added level of randomness in the progression of scale parameters: as before (but specified in reverse order), we assume apriori that $d \sim \Gamma(a_{I+1}, y)$, that for $1 \leq i \leq I$ the terms c_i are sampled from independent $\Gamma(a_i, d)$ distributions, that for each given i and $1 \leq l \leq L$ the terms $b_{i,l} \sim \Gamma(\bar{a}_{i,l}, c_i)$ are sampled independently and that lastly for $1 \leq j \leq J$ and given i, l the data points $\{x_{i,l,j}\}$ are assumed to have

been sampled from independent $\Gamma(\tilde{a}_{i,l,j}, b_{i,l})$ distributions. The posterior density function, as well as the marginal density functions can be easily derived and the result is that a sequentially updating Gibbs sampler (similar to what we have in (2.14)) would work as follows:

$$\begin{aligned}
 b_{i,l}^{t+1} &= \frac{\gamma_{i,l}^{t+1}}{\sum_{j=1}^J x_{i,l,j} + c_i^t} \\
 c_i^{t+1} &= \frac{\gamma_i^{t+1}}{\sum_{l=1}^L b_{i,l}^{t+1} + d^t} \\
 d^{t+1} &= \frac{\gamma_{I+1}^{t+1}}{\sum_{i=1}^I c_i^{t+1} + y}
 \end{aligned} \tag{2.18}$$

This random dynamical system is no longer reducible to a simplified 1-dimensional system. The same follows for any similar variation (for example altering the order in which the co-ordinates are renewed), and the method used in the previous sections is no longer viable.

3 An attempt at generalisation

What is evident from the proof of Theorem 2.2 (and hence by similarity also from the proof of Theorem 2.6) is its dependence on certain ad hoc properties of this particular problem (mainly relating to the gamma distribution). In this section we will attempt to provide some condition that would guarantee that a random iterated function system $\{f_t\}$ is 1-contractive 'in the long run'.

Let $\{f_t\}$ be non-decreasing, i.i.d. random functions with domain $D \in \mathbb{R}$ and range $R \subseteq D \cap (0, \infty)$. Suppose that there is some $\zeta \in D$ and $C > 1$ such that $f_t(y) \geq Cy$ for all t and for all $y \in D$, $y \leq \zeta$. Suppose also that $\mathbb{E}[Lip(f_1)] \leq r < 1$ on $D \cap (\zeta, \infty)$, where $Lip(f_1)$ is the Lipschitz constant of the random function f_1 on $D \cap (\zeta, \infty)$.

Theorem 3.1. *For the iterated function system $\{f_t\}$ satisfying the above conditions, there is a r.v. F_∞ such that for all x , $F_t(x) \rightarrow F_\infty$ a.s. Furthermore, if $x \geq \zeta$*

$$\mathbb{E} \left[|F_{t+1}(x) - F_\infty| \right] \leq \frac{r^t}{1-r} \mathbb{E} \left[|f_1(x) - x| \right]$$

and if $x < \zeta$,

$$\mathbb{E} \left[|F_{t+\tau}(x) - F_\infty| \right] \leq \frac{r^t}{1-r} \mathbb{E} \left[|f_1 \circ \dots \circ f_\tau(x) - f_{\tau+1} \circ f_1 \circ \dots \circ f_\tau(x)| \right] \quad (3.1)$$

where $\tau = \left\lceil \ln \left(\frac{\zeta}{x} \right) / \ln(C) \right\rceil$.

The expectation in the right-hand side of (3.1) can often be easily bounded, given the distribution of the functions f_t .

Proof. Since f_1 is non-decreasing, if $x \geq \zeta$ then $f_1(x) \geq f_1(\zeta) \geq \zeta$. Then existence of F_∞ follows from Theorem 2.1 (see also paragraph preceding Theorem 2.1), as does the bound for the case $x \geq \zeta$. If $x < \zeta$, note that since $f_i(x) \geq Cx$, we get $f_{t+j} \circ \dots \circ f_{t+\tau}(x) \geq \zeta$ for some $1 \leq j \leq \tau$. By the previous remark, we have that

$$f_{t+1} \circ \dots \circ f_{t+j} \circ \dots \circ f_{t+\tau}(x) \geq \zeta$$

Then

$$\begin{aligned} \mathbb{E} \left[|F_{t+\tau}(x) - F_\infty| \right] &= \mathbb{E} \left[\left| F_t(f_{t+1} \circ \dots \circ f_{t+\tau}(x)) - F_\infty \right| \right] \\ &\leq \frac{r^t}{1-r} \mathbb{E} \left[\left| f_1 \circ \dots \circ f_\tau(x) - f_{\tau+1}(f_1 \circ \dots \circ f_\tau(x)) \right| \right] \end{aligned}$$

where the second inequality follows from the previous case (since $f_1 \circ \dots \circ f_\tau(x) \geq \zeta$). □

To illustrate how this result can be applied, we can consider as an example the following logistic-type random dynamical system

$$f_t(x) := \left(\frac{\alpha_t}{1 + \delta x} + \beta_t \right) x \quad (3.2)$$

Here α_t and β_t are non-negative, real-valued random variables (with finite first moments) representing growth and decay factors, respectively. We will require that $\mathbb{E}[\beta_t] = \rho < 1$ and that $\alpha_t \geq a_m$ for some $a_m \in \mathbb{R}_+$. The term δ provides a degree of freedom in stating for which values of x the function f_t is decaying, and for which it is growing. We confirm that $f'_t(x) = \beta_t + \frac{\alpha_t}{(1+\delta x)^2} \geq 0$ and $\mathbb{E}[Lip(f_t)] \leq r := \frac{\rho+1}{2}$ in the domain $\left(\frac{1}{\delta} \left(\sqrt{\frac{2\mathbb{E}[\alpha_t]}{1-\rho}} - 1 \right), \infty \right)$.

We would like to assert the condition that $f_t(x) \geq Cx$ for $x \leq \frac{1}{\delta} \left(\sqrt{\frac{2\mathbb{E}[\alpha_t]}{1-\rho}} - 1 \right)$, for some $C > 1$. For such values of x , this is equivalent to $\frac{\alpha_t}{1+\delta x} + \beta_t \geq C$, and a sufficient condition would be that $a_m \left(1 + \delta \cdot \frac{1}{\delta} \left(\sqrt{\frac{2\mathbb{E}[\alpha_t]}{1-\rho}} - 1 \right) \right)^{-1} > 1$, or in other words $a_m > \sqrt{\frac{2\mathbb{E}[\alpha_t]}{1-\rho}}$. Then we can take $C = a_m \left(\frac{2\mathbb{E}[\alpha_t]}{1-\rho} \right)^{-\frac{1}{2}}$, and by the previous theorem

$$\mathbb{E} \left[|F_{t+1}(x) - F_\infty| \right] \leq \frac{r^t}{1-r} \mathbb{E} \left[|f_{t+1}(x) - x| \right]$$

for $x \geq \zeta := \frac{1}{\delta} \left(\sqrt{\frac{2\mathbb{E}[\alpha_t]}{1-\rho}} - 1 \right)$, and

$$\mathbb{E} \left[|F_{t+\tau}(x) - F_\infty| \right] \leq \frac{r^t}{1-r} \mathbb{E} \left[\left| f_1 \circ \dots \circ f_\tau(x) - f_{\tau+1}(f_1 \circ \dots \circ f_\tau(x)) \right| \right] \quad (3.3)$$

for $x \in (0, \zeta)$ and $\tau := \left\lceil \ln \left(\frac{\zeta}{x} \right) / \ln(C) \right\rceil$.

A simple bound for the term on the right-hand side of (3.3) can be obtained as follows: let $y := f_1 \circ \dots \circ f_\tau(x)$, and observe also that $f_t(x) \leq \frac{\alpha_t}{\delta} + \beta_t x$. Hence

$$\begin{aligned}
& \mathbb{E} \left[\left| f_1 \circ \dots \circ f_\tau(x) - f_{\tau+1}(f_1 \circ \dots \circ f_\tau(x)) \right| \right] \\
&= \mathbb{E} \left[y \left| \frac{\alpha_{\tau+1}}{1 + \delta y} + \beta_{\tau+1} - 1 \right| \right] \\
&\leq \mathbb{E} \left[y \cdot \frac{\alpha_{\tau+1}}{1 + \delta \zeta} \right] + \mathbb{E} [y |\beta_{\tau+1} - 1|] \\
&\leq \mathbb{E} [y] \left(\sqrt{\frac{\mathbb{E}[\alpha_t](1 - \rho)}{2}} + 1 + \rho \right) \\
&\leq \left(\sqrt{\frac{\mathbb{E}[\alpha_t](1 - \rho)}{2}} + 1 + \rho \right) \left(\frac{\mathbb{E}[\alpha_t]}{\delta} (1 + \rho + \dots + \rho^{\tau-1}) + \rho^\tau x \right) \\
&\leq \left(\sqrt{\frac{\mathbb{E}[\alpha_t](1 - \rho)}{2}} + 1 + \rho \right) \left(\frac{\mathbb{E}[\alpha_t]}{\delta(1 - \rho)} + \rho^\tau x \right)
\end{aligned}$$

To make this example more concrete, let us assume that $\{\beta_t\}$ are i.i.d uniform on $[0, 1]$, $\{\alpha_t\}$ are i.i.d uniform on $[5, 6]$, and $\delta = 1$. Then

$$\mathbb{E} \left[|F_{t+1}(x) - F_\infty| \right] \leq 4 \left(\frac{3}{4} \right)^t \left| \frac{x}{2} - 5.5 \right|$$

for $x \geq 4.7$, and

$$\mathbb{E} \left[|F_{t+\tau}(x) - F_\infty| \right] \leq 4 \left(\frac{3}{4} \right)^t \left| 2.7 \left(11 + \left(\frac{1}{2} \right)^\tau x \right) \right| \quad (3.4)$$

for $x \in (0, 4.7)$ and $\tau := \left\lceil 15.7 \ln \left(\frac{3.7}{x} \right) \right\rceil$.

This example is not a 1-contractive random dynamical system (which can be concluded from a quick inspection of f'_1), hence a direct application of Theorem 2.1

cannot be done. However, by Theorem 3.1 we were able to infer an explicit bound corresponding to a geometric rate of convergence.

4 A different approach

4.1 The problem

We will return to our original problem of constructing a rapidly mixing Markov chain on \mathbb{R}_+^n that converges to the target distribution with density function given by (2.3).

We will first consider the Markov chain which sequentially updates its co-ordinates as follows: for $i \in \{1, 2, \dots, n\}$ let

$$\bar{P}_i(v, dw) := \left(\prod_{j \neq i} \delta_{v_j}(w_j) \right) h_i(w_i | v) dw_i$$

where $h_i(w_i | v)$ is the density function of $\frac{\gamma_i}{v_{i-1} + v_{i+1}}$ given v , and where for convenience we have defined $v_0 := x$ and $v_{n+1} := b$. Let

$$\bar{P} := \bar{P}_1 \bar{P}_3 \cdots \bar{P}_{2 \lfloor \frac{n}{2} \rfloor - 1} \bar{P}_2 \bar{P}_4 \cdots \bar{P}_{2 \lfloor \frac{n}{2} \rfloor} \quad (4.1)$$

or in other words \bar{P} updates all odd-numbered co-ordinates first, followed by the even numbered co-ordinates. For the case $n = 4$, this will be the main algorithm we will study, and we will work with different algorithms for the $n > 4$, which

will be defined later. We can also motivate the construction of (4.1) with the following map which may be thought of as a “simultaneous” Gibbs sampler: for $a_1, a_2, \dots, a_{n+1} \in \mathbb{R}_+$ and $\gamma_i \sim \Gamma(a_i + a_{i+1}, 1)$ independent, we define

$$(u_1, u_2, \dots, u_n) \longrightarrow \left(\frac{\gamma_1}{x + u_2}, \frac{\gamma_2}{u_1 + u_3}, \dots, \frac{\gamma_n}{u_{n-1} + b} \right) \quad (4.2)$$

where $x, b \in \mathbb{R}_+$. An immediate consequence is that for $t > 1$ we can eliminate the odd numbered co-ordinates by considering the 2-step (sub)Markov chain derived from (4.2). If n is odd this becomes

$$\begin{aligned} & (u_2^{t+1}, u_4^{t+1}, \dots, u_{n-1}^{t+1}) \\ &= \left(\frac{\gamma_2^{t+1}}{\frac{\gamma_1^t}{x+u_2^{t-1}} + \frac{\gamma_3^t}{u_2^{t-1}+u_4^{t-1}}}, \frac{\gamma_4^{t+1}}{\frac{\gamma_3^t}{u_2^{t-1}+u_4^{t-1}} + \frac{\gamma_5^t}{u_4^{t-1}+u_6^{t-1}}}, \dots, \frac{\gamma_{n-1}^{t+1}}{\frac{\gamma_{n-2}^t}{u_{n-3}^{t-1}+u_{n-1}^{t-1}} + \frac{\gamma_n^t}{u_{n-1}^{t-1}+b}} \right) \end{aligned}$$

and similarly, for n even this would only differ in the final co-ordinate,

$$\begin{aligned} & (u_2^{t+1}, u_4^{t+1}, \dots, u_n^{t+1}) \\ &= \left(\frac{\gamma_2^{t+1}}{\frac{\gamma_1^t}{x+u_2^{t-1}} + \frac{\gamma_3^t}{u_2^{t-1}+u_4^{t-1}}}, \frac{\gamma_4^{t+1}}{\frac{\gamma_3^t}{u_2^{t-1}+u_4^{t-1}} + \frac{\gamma_5^t}{u_4^{t-1}+u_6^{t-1}}}, \dots, \frac{\gamma_n^{t+1}}{\frac{\gamma_{n-1}^t}{u_{n-2}^{t-1}+u_n^{t-1}} + b} \right) \end{aligned}$$

Therefore, if we work with the Markov chain on \mathbb{R}_+^n defined by

$$\begin{aligned} u_{2i-1}^{t+1} &= \frac{\gamma_{2i-1}^{t+1}}{u_{2i-2}^t + u_{2i}^t} \\ u_{2i}^{t+1} &= \frac{\gamma_{2i}^{t+1}}{u_{2i-1}^{t+1} + u_{2i+1}^{t+1}} \end{aligned} \quad (4.3)$$

for $1 \leq i \leq \lfloor \frac{n}{2} \rfloor$ (where $u_0^t = x$ and $u_{n+1}^t = b$ for all t), by Corollary 4.1 we can reduce the dimensions of the problem by considering the Markov chain defined by a random function on $\mathbb{R}_+^{\lfloor \frac{n}{2} \rfloor}$ which for n odd is given by:

$$f_n(u_2, u_4, \dots, u_{n-1}) = \left(\frac{\gamma_2}{\frac{\gamma_1}{x+u_2} + \frac{\gamma_3}{u_2+u_4}}, \frac{\gamma_4}{\frac{\gamma_3}{u_2+u_4} + \frac{\gamma_5}{u_4+u_6}}, \dots, \frac{\gamma_{n-1}}{\frac{\gamma_{n-2}}{u_{n-3}+u_{n-1}} + \frac{\gamma_n}{u_{n-1}+b}} \right) \quad (4.4)$$

A similar definition results for the case where n is even. Note that if Y^t, U^t are instances of the Markov chains defined by (4.2) and (4.4) respectively, then $U^t \sim \left(Y_2^{2t}, Y_4^{2t}, \dots, Y_{2\lfloor \frac{n}{2} \rfloor}^{2t} \right)$.

Corollary 4.1. *Let Υ^t be a copy of the Markov chain (4.3) on \mathbb{R}_+^n and let Φ^t be a copy of (4.4) on $\mathbb{R}_+^{\lfloor \frac{n}{2} \rfloor}$, and let $\bar{\pi}$ and π be their respective equilibrium distributions. Then $d_{TV}(\Upsilon^{t+1}, \bar{\pi}) \leq d_{TV}(\Phi^t, \pi)$.*

Proof. If we use the notation $\Phi^t = \left(\Phi_2^t, \Phi_4^t, \dots, \Phi_{2\lfloor \frac{n}{2} \rfloor}^t \right)$, then $\left(\Upsilon_2^t, \Upsilon_4^t, \dots, \Upsilon_{2\lfloor \frac{n}{2} \rfloor}^t \right) \sim \Phi^t$. There are random variables $\Psi = \left(\Psi_2, \Psi_4, \dots, \Psi_{2\lfloor \frac{n}{2} \rfloor} \right) \sim \Phi^t$ and $\Lambda = \left(\Lambda_2, \Lambda_4, \dots, \Lambda_{2\lfloor \frac{n}{2} \rfloor} \right) \sim \pi$ such that

$$d_{TV}(\Phi^t, \pi) = \mathbb{P}[\Psi \neq \Lambda]$$

But since π is the marginal distribution of all even co-ordinates of $\bar{\pi}$, we have that

$$\bar{\pi} \sim \Xi := \left(\frac{\gamma_1^{t+1}}{x + \Lambda_2}, \frac{\gamma_2^{t+1}}{\frac{\gamma_1^{t+1}}{x+\Lambda_2} + \frac{\gamma_3^{t+1}}{\Lambda_2+\Lambda_4}}, \frac{\gamma_3^{t+1}}{\Lambda_2 + \Lambda_4}, \dots \right)$$

and similarly

$$\Upsilon^{t+1} \sim \Theta := \left(\frac{\gamma_1^{t+1}}{x + \Psi_2}, \frac{\gamma_2^{t+1}}{\frac{\gamma_1^{t+1}}{x + \Psi_2} + \frac{\gamma_3^{t+1}}{\Psi_2 + \Psi_4}}, \frac{\gamma_3^{t+1}}{\Psi_2 + \Psi_4}, \dots \right)$$

Hence

$$\begin{aligned} d_{TV}(\Upsilon^{t+1}, \bar{\pi}) &\leq \mathbb{P}[\Xi \neq \Theta] \\ &\leq \mathbb{P}[\Psi \neq \Lambda] \\ &= d_{TV}(\Phi^t, \pi) \end{aligned}$$

□

Now if $\bar{\pi}$ is the probability measure on \mathbb{R}_+^n with density function (2.3), it is a well known fact (e.g. see Section 2.3 of [27]) that $\bar{\pi}$ is the equilibrium distribution of the Markov chain defined by (4.3). It follows that the marginal distribution of the even co-ordinates of $\bar{\pi}$, which we denote by π , is the equilibrium distribution of (4.4).

We can now state our main results. For $n = 4$, let \mathcal{U}^t and \mathcal{V}^t be two copies of the Markov chain starting at points \mathcal{U}^0 and \mathcal{V}^0 respectively. Let d_{TV} denote the total variation metric on probability measures of a probability space Ω , defined by

$$d_{TV}(\nu_1, \nu_2) := \sup_{A \subseteq \Omega} |\nu_1(A) - \nu_2(A)| \quad (4.5)$$

For two random variables $X \sim \nu_1$ and $Y \sim \nu_2$, we let $d_{TV}(X, Y) := d_{TV}(\nu_1, \nu_2)$.

We will define the condition

$$a_1 + a_4 > 1, a_2 + a_5 > 1, a_2 + a_3 > 1, a_3 + a_4 > 1, a_4 + a_5 > 1 \quad (4.6)$$

and relating to the statement of Theorem 4.2, we will also define the following

terms:

$M = \max_i \{\mathcal{U}_i^0, \mathcal{V}_i^0\}, m = \min_i \{\mathcal{U}_i^0, \mathcal{V}_i^0\}$
$R_0 = \frac{M}{m}, J_0 = 2m + m(3M + b)$
$C_1 := \frac{a_2 + a_3}{a_1 + a_2 + a_3 + a_4 - 1}x + \frac{a_4 + a_5}{b}, C_2 := \frac{a_1 + a_2 + xb}{x(a_2 + a_3 + a_4 + a_5 - 1)}$
$\eta = \frac{C_1 + C_2}{1 - \max\{(a_2 + a_3)/(a_1 + a_2 + a_3 + a_4 - 1), (a_3 + a_4)/(a_2 + a_3 + a_4 + a_5 - 1)\}}$
$\varsigma = \frac{4(a_3 + a_4)}{(a_1 + a_2 - \frac{1}{3})} + 4$
$\theta_1 = \frac{\varsigma}{x} \frac{a_2 + a_3}{a_1 + a_2 + a_3 + a_4 - 1}$
$\theta_2 = \mathbb{E} \left[\left(2 + \frac{\gamma_2}{\gamma_4} + \frac{\gamma_4}{\gamma_2} \right) \left(\frac{\gamma_3}{\gamma_2 + \gamma_4} \right) \right] \left(\varsigma x + \frac{4(a_3 + a_4)}{b} \right)$
$\theta_3 = \frac{\varsigma}{x} \left(\frac{a_2 + a_3}{a_1 + a_2 + a_3 + a_4 - 1}x + \frac{a_4 + a_5}{b} \right) + \left(\varsigma x + \frac{a_4 + a_5}{b} \right) \mathbb{E} \left[\left(2 + \frac{\gamma_2}{\gamma_4} + \frac{\gamma_4}{\gamma_2} \right) \left(\frac{\frac{\gamma_1}{x} + b}{\gamma_2 + \gamma_4} \right) \right]$
$r = 1 - (\eta(\theta_1 + \theta_2) + \theta_3)^{-1}$
$\beta = \frac{1 + \max\{(a_2 + a_3)/(a_1 + a_2 + a_3 + a_4 - 1), (a_3 + a_4)/(a_2 + a_3 + a_4 + a_5 - 1)\}}{2}$
$d' := \ln(\beta) ^{-1} \left(2 - \frac{\ln(r)}{2} \right) + 2$

Then

Theorem 4.2. [Case $n = 4$] Suppose that (4.6) holds. If $J_0 \leq \eta$, then for $t \geq d'$,

$$d_{TV}(\mathcal{U}^{t+3}, \mathcal{V}^{t+3}) \leq r^{\frac{1}{2} \lfloor \frac{t}{d'} \rfloor} (1 + 3(a_2 + a_3 + a_4 + a_5)(R_0 - 1))$$

For general values of J_0 and $t \geq d'$, we have that

$$d_{TV}(\mathcal{U}^{t+3}, \mathcal{V}^{t+3}) \leq r^{\frac{1}{2} \lfloor \frac{t}{2d'} \rfloor} (1 + 3(a_2 + a_3 + a_4 + a_5)(R_0 - 1)) + \frac{\max\{J_0, \eta\}}{\eta} \beta^{\lfloor \frac{t}{2} \rfloor + 3}$$

Note that the definition of θ_2 may be amended to contain an upper bound on the expectation (e.g. $\frac{2(a_3+a_4)}{a_2+a_3+a_4+a_5-1} + \frac{a_3+a_4}{a_4+a_5-1} + \frac{a_3+a_4}{a_2+a_3-1}$) rather than an exact value. The same can be done with the definition of θ_3 . Our need for condition (4.6) becomes clear in section 4.7. If we let $\mathcal{U}^0 = (1, 1, 1, 1)$ and $\mathcal{V}^0 \sim \pi$ then

Corollary 4.3. [Case $n = 4$] For $t \geq d'$,

$$d_{TV}(\mathcal{U}^{t+3}, \pi) \leq \mathbb{E}_\pi[R_0] r^{\frac{1}{2} \lfloor \frac{t}{2d'} \rfloor} (1 + 3(a_2 + a_3 + a_4 + a_5)) + \left(\frac{\mathbb{E}_\pi[J_0]}{\eta} + 1 \right) \beta^{\lfloor \frac{t}{2} \rfloor + 3}$$

The quantities $\mathbb{E}_\pi[R_0]$ and $\mathbb{E}_\pi[J_0]$ depend only on π and can be estimated easily.

For $n = 2m > 4$ (we make an observation later that the analysis for odd n is nearly identical), we let \mathcal{U}^t and \mathcal{V}^t evolve according to a slightly different Gibbs sampler, in which coordinates are updated in numerical order. Then

Theorem 4.4. [Case n even, $n > 4$] For the Markov chain given in (4.43), suppose that $\max\{\zeta_2, \dots, \zeta_{2m}, \xi_2, \dots, \xi_{2m}\} < 1$. Then

$$d_{TV}(\mathcal{U}^{t+3}, \mathcal{V}^{t+3}) \leq r^{\lfloor \frac{t}{2d'} \rfloor} (1 + 3(a_2 + a_3 + \dots + a_{2m+1})(R_0 - 1)) + \frac{\max\{J_0, \eta\}}{\eta} \beta^{\lfloor \frac{t}{2} \rfloor + 3}$$

Here the terms $\zeta_2, \dots, \zeta_{2m}$ and ξ_2, \dots, ξ_{2m} are defined in (4.55) and (4.58) respectively, and depend only on the parameters x, b and $\{a_i\}$. $R_0 = \frac{M}{\varrho}$ and $J_0 = m\varrho + \frac{m}{\varrho}$, where $M = \max_i \{\mathcal{U}_i^0, \mathcal{V}_i^0\}$, $\varrho = \min_i \{\mathcal{U}_i^0, \mathcal{V}_i^0\}$, while the terms $r < 1$, d' , $\beta < 1$ and η are defined in Section 4.11.

4.2 Outline of our proof

Essentially the proof of Theorem 4.2 and Theorem 4.4 are quite similar, both relying on a coupling argument. In Section 4.3 we define a partial order ' \preceq ' on $\mathbb{R}_+^{\lfloor \frac{n}{2} \rfloor}$ and show that we can couple two copies $\{u^t, v^t\}$ of (4.4), with the initial condition $u^0 \preceq v^0$, in a stochastically monotone manner, thus preserving the order $u^t \preceq v^t$ for all time t . In the beginning of Section 4.4 we show that if R_t is a process that serves as an upper bound for the ratio $\max_i \left\{ \frac{v_i^t}{u_i^t} \right\}$, then the rate of convergence of $R_t \rightarrow 1$ can be related to the rate at which (4.4) converges to equilibrium. Therefore, our focus becomes the defining of such a process and showing that it converges to 1 at a geometric rate.

We define v^t in Section 4.5 (for the case $n = 4$) to be a stochastic process adapted to the same filtration as u^t , with the property that it is an upper bound to (in the sense of \preceq) a faithful copy of (4.4) started at v^0 . This also allows us to give an exact definition of the previously mentioned process R_t , which then also proves to have the additional quality of being strictly monotone decreasing. This alone does not guarantee that $R_t \rightarrow 1$ quickly (or at any pace, for that matter). 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 $R_t \rightarrow 1$ at a geometric rate. To fulfill this condition, we define a

number of auxiliary processes in Section 4.6 (and show their existence in Section 4.7) that serve as an upper bound for the terms $\left\{u_2^t, u_4^t, \frac{1}{u_2^t}, \frac{1}{u_4^t}\right\}$, and we show that they are frequently bounded from above by some constant η .

The case $n > 4$ is treated in Section 4.10. We define a Markov chain somewhat different from (4.4), for the purpose of obtaining a monotone decreasing process R_t that has the desired properties mentioned above. The proof of Theorem 4.4 follows in an analogous manner to what we have for Theorem 4.2, however finding the required auxiliary processes proves to be more elusive. We show their existence under certain constraints on the parameters.

4.3 Stochastically monotone coupling

For $u, v \in \mathbb{R}_+^n$, define the partial order $u \preceq v$ to mean $u_i \leq v_i$ for even i , and $u_i \geq v_i$ for odd i . For the 'reduced' chain (4.4) on $\mathbb{R}_+^{\lfloor \frac{n}{2} \rfloor}$ we can take this partial order to imply the same (since we are only concerned with the even co-ordinates, this would mean that we have pointwise inequality in the same direction at every co-ordinate).

Suppose we couple two copies of (4.2), $u^0 \preceq v^0$, by employing the same random variables $\{\gamma_i^t\}$ in both copies (we will refer to this as the 'uniform coupling'). Then $u^t \preceq v^t$ for all times t . Therefore if we couple in this manner two copies commencing at some arbitrary initial points $\mathcal{U}^0, \mathcal{V}^0 \in \mathbb{R}_+^n$, we can take $m = \min \{\mathcal{U}_1^0, \dots, \mathcal{U}_n^0, \mathcal{V}_1^0, \dots, \mathcal{V}_n^0\}$ and $M = \max \{\mathcal{U}_1^0, \dots, \mathcal{U}_n^0, \mathcal{V}_1^0, \dots, \mathcal{V}_n^0\}$,

and define

$$\begin{aligned} v^0 &:= (m, M, m, \dots) \in \mathbb{R}_+^n \\ u^0 &:= (M, m, M, \dots) \in \mathbb{R}_+^n \end{aligned} \tag{4.7}$$

i.e. we are setting $v_{2j+1}^0 = u_{2j+2}^0 = m$ and $v_{2j+2}^0 = u_{2j+1}^0 = M$. And by observing that $u^0 \preceq \{\mathcal{U}^0, \mathcal{V}^0\} \preceq v^0$, we conclude that \mathcal{U}^t and \mathcal{V}^t are perpetually 'squeezed' between u^t and v^t (or in other words $u^t \preceq \{\mathcal{U}^t, \mathcal{V}^t\} \preceq v^t$). We can justify with Corollary 4.6 why it suffices to consider the coupled pair (u^t, v^t) in order to bound $d_{TV}(\mathcal{U}^t, \mathcal{V}^t)$.

Lemma 4.5. *Suppose that $0 < \beta_1 < \beta_2 < \beta_3 < \beta_4$, and let $z_i \sim \Gamma(\alpha, \beta_i)$ Then*
 $d_{TV}(z_2, z_3) \leq d_{TV}(z_1, z_4)$

Proof. Let f_1, f_2, f_3, f_4 be the respective density functions. By the following property of total variation (see Theorem 5.7 of [1])

$$d_{TV}(z_i, z_j) = 1 - \int \min(f_i(y), f_j(y)) dy$$

it is enough to show that $\min\{f_1(y), f_4(y)\} \leq \min\{f_2(y), f_3(y)\}$ for all y . Note first that for $i, j \in \{1, 2, 3, 4\}$ with $i < j$,

$$\begin{aligned} f_i(y) &\geq f_j(y) \\ \iff \beta_i^\alpha \exp(-\beta_i y) &\geq \beta_j^\alpha \exp(-\beta_j y) \\ \iff y &\geq \frac{\alpha (\ln(\beta_j) - \ln(\beta_i))}{\beta_j - \beta_i} \end{aligned} \tag{4.8}$$

Let

$$g(\beta, \kappa) := \frac{\alpha (\ln(\beta) - \ln(\kappa))}{\beta - \kappa}$$

then

$$\frac{\partial g}{\partial \kappa} = \frac{\alpha(1 - \frac{\beta}{\kappa} + \ln(\frac{\beta}{\kappa}))}{(\beta - \kappa)^2}$$

The numerator of this equation is non-positive for all $\beta, \kappa \in \mathbb{R}_+$. This can be seen by observing that the function $\ln(z) - z$ achieves a global maximum on $(0, \infty)$ at $z = 1$ with value -1 . Hence $1 - \frac{\beta}{\kappa} + \ln(\frac{\beta}{\kappa}) \leq 0$, and since $g(\beta, \kappa)$ is non-increasing in κ , we have the following relation:

$$g(\beta_4, \beta_3) \leq g(\beta_4, \beta_2) \leq g(\beta_4, \beta_1) = g(\beta_1, \beta_4) \leq g(\beta_1, \beta_3) \leq g(\beta_1, \beta_2) \quad (4.9)$$

Then from (4.8) and (4.9) it follows that

$$\begin{aligned} f_1(y) &\leq \min \{f_2(y), f_3(y)\} \text{ on } [0, g(\beta_1, \beta_3)] \\ f_4(y) &\leq \min \{f_2(y), f_3(y)\} \text{ on } [g(\beta_4, \beta_2), \infty) \end{aligned}$$

hence

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

□

Let $\mathcal{U}_{\mathcal{U}^t}^{t+1}$ be a random vector having the conditional distribution of \mathcal{U}^{t+1} given \mathcal{U}^t , with similar definitions for $\mathcal{V}_{\mathcal{V}^t}^{t+1}$, $u_{\mathcal{U}^t}^{t+1}$ and $v_{\mathcal{V}^t}^{t+1}$. We apply the uniform coupling until

time t , and given this outcome we couple $(\mathcal{U}_{|\mathcal{U}^t}^{t+1}, \mathcal{V}_{|\mathcal{V}^t}^{t+1}, u_{|u^t}^{t+1}, v_{|v^t}^{t+1})$ in the following “one-shot” manner (described in [10] in further detail): for each co-ordinate i , we take $u_{|u^t}^{t+1}$ to be the x-coordinate of a uniformly chosen point from the area under the graph of the density function f_{u_i} of $u_{|u^t}^{t+1}$. Set $v_{|v^t}^{t+1} = \mathcal{V}_{|\mathcal{V}^t}^{t+1} = \mathcal{U}_{|\mathcal{U}^t}^{t+1} = u_{|u^t}^{t+1}$ if this point also lies below the graph of the density function f_{v_i} of $v_{|v^t}^{t+1}$. Otherwise, take $v_{|v^t}^{t+1}$ to be the x-coordinate of a uniformly and independently chosen point from the area above the graph of $\min\{f_{u_i}, f_{v_i}\}$ and below the graph of f_{v_i} . Similarly, set $\mathcal{V}_{|\mathcal{V}^t}^{t+1}$ to be the x-coordinate of a uniformly and independently chosen point from the area above the graph of $\min\{f_{u_i}, f_{v_i}\}$ and below the graph of f_{v_i} , and set $\mathcal{U}_{|\mathcal{U}^t}^{t+1}$ to be the x-coordinate of a uniformly and independently chosen point from the area above the graph of $\min\{f_{u_i}, f_{v_i}\}$ and below the graph of f_{u_i} . From the proof of Lemma 4.5 it can be seen that $\min\{f_{u_i}, f_{v_i}\} \leq \min\{f_{v_i}, f_{u_i}\}$, hence it is easy to verify that this is indeed a coupling of $(\mathcal{U}_{|\mathcal{U}^t}^{t+1}, \mathcal{V}_{|\mathcal{V}^t}^{t+1}, u_{|u^t}^{t+1}, v_{|v^t}^{t+1})$.

Corollary 4.6. *With one-shot coupling at time $t + 1$, we have $d_{TV}(\mathcal{U}^{t+1}, \mathcal{V}^{t+1}) \leq \mathbb{P}[u^{t+1} \neq v^{t+1}]$*

Proof. We will first observe the following

$$\begin{aligned}
d_{TV} \left(\mathcal{U}_{|\mathcal{U}^t}^{t+1}, \mathcal{V}_{|\mathcal{V}^t}^{t+1} \right) &\leq \mathbb{P} \left[\mathcal{U}_{|\mathcal{U}^t}^{t+1} \neq \mathcal{V}_{|\mathcal{V}^t}^{t+1} \right] \\
&= \mathbb{P} \left[\cup_i \left\{ \mathcal{U}_{i,|\mathcal{U}^t}^{t+1} \neq \mathcal{V}_{i,|\mathcal{V}^t}^{t+1} \right\} \right] \\
&\leq \mathbb{P} \left[\cup_i \left\{ u_{i,|\mathcal{U}^t}^{t+1} \neq v_{i,|\mathcal{V}^t}^{t+1} \right\} \right] \\
&= \mathbb{P} \left[u_{|\mathcal{U}^t}^{t+1} \neq v_{|\mathcal{V}^t}^{t+1} \right]
\end{aligned}$$

The inequality in the third line is a consequence of the previous lemma and the fact that $\min \{f_{\mathcal{U}_i}, f_{\mathcal{V}_i}\} \geq \min \{f_{u_i}, f_{v_i}\}$ implies

$\left\{ \mathcal{U}_{i,|\mathcal{U}^{t-1}}^t \neq \mathcal{V}_{i,|\mathcal{V}^{t-1}}^t \right\} \subseteq \left\{ u_{i,|\mathcal{U}^{t-1}}^t \neq v_{i,|\mathcal{V}^{t-1}}^t \right\}$. This inequality holds for any outcome of $(\mathcal{U}^t, \mathcal{V}^t, u^t, v^t)$ so long as the partial order $u^t \preceq \{\mathcal{U}^t, \mathcal{V}^t\} \preceq v^t$ persists. But by the previous remarks this is always the case for the uniform coupling, hence the statement of the corollary. \square

4.4 The ratio R_t

From here on we will mainly consider the 'reduced' Markov chain defined by (4.4), however at times it will be useful to refer to odd numbered co-ordinates as derived from the transition kernel \bar{P} in (4.1).

We assume in this section that $u^t = \left(u_2^t, u_4^t, \dots, u_{2 \lfloor \frac{n}{2} \rfloor}^t \right) \preceq \left(v_2^t, v_4^t, \dots, v_{2 \lfloor \frac{n}{2} \rfloor}^t \right) = v^t$, so that $\frac{v_i^t}{u_i^t} \geq 1$. Define the filtration $\mathcal{F}_t := \sigma(u^0, v^0, \gamma_1^1, \dots, \gamma_n^1, \dots, \gamma_1^t, \dots, \gamma_n^t)$ and let R_t be a non-increasing \mathcal{F}_t -measurable process such that $R_t \geq \max_i \left\{ \frac{v_i^t}{u_i^t} \right\}$.

Then $u^t = v^t$ if $R_t = 1$. The uniform coupling defined for (4.2) in the previous section can also be easily applied to (4.4) as well as the result of Corollary 4.6 (where in the final step at time $t+1$ one would couple $\left(\gamma_1^{t+1}, \gamma_3^{t+1}, \dots, \gamma_{2\lceil \frac{n}{2} \rceil - 1}^{t+1}\right)$ uniformly, while taking the 'one shot' approach for $\left(\gamma_2^{t+1}, \gamma_4^{t+1}, \dots, \gamma_{2\lfloor \frac{n}{2} \rfloor}^{t+1}\right)$ as described in the section preceding Corollary 4.6). Under these assumptions we can now relate R_t to $\mathbb{P}[u^{t+1} \neq v^{t+1} | \mathcal{F}_t]$.

Lemma 4.7. *Applying one-shot coupling at time $t+1$, we have*

$$\mathbb{P}[u^{t+1} \neq v^{t+1} | \mathcal{F}_t] \leq 1 - R_t^{-\sum_{i=1}^{\lfloor \frac{n}{2} \rfloor} (a_{2i} + a_{2i+1})}$$

Proof. Let $h_{u_{2i}}(y | \mathcal{F}_t, \gamma_{2i-1}^{t+1}, \gamma_{2i+1}^{t+1})$ and $h_{v_{2i}}(y | \mathcal{F}_t, \gamma_{2i-1}^{t+1}, \gamma_{2i+1}^{t+1})$ be the conditional density functions of u_{2i}^{t+1} and v_{2i}^{t+1} given

$\mathcal{F}_t, \gamma_{2i-1}^{t+1}, \gamma_{2i+1}^{t+1}$ (and therefore, also given $u_{2i-1}^{t+1}, u_{2i+1}^{t+1}, v_{2i-1}^{t+1}, v_{2i+1}^{t+1}$ where $u_{2i-1}^{t+1} = \gamma_{2i-1}^{t+1} / (u_{2i-2}^t + u_{2i-2}^t)$ and we define $u_{2i+1}^{t+1}, v_{2i-1}^{t+1}$ and v_{2i+1}^{t+1} similarly). These represent gamma random variables with shape parameters given by $a_{2i} + a_{2i+1}$ and scale parameters $u_{2i-1}^{t+1} + u_{2i+1}^{t+1}$ and $v_{2i-1}^{t+1} + v_{2i+1}^{t+1}$ respectively, as can be seen from the definition of the transition kernel \bar{P} . Then

$$h_{u_{2i}}(y | \mathcal{F}_t, \gamma_{2i-1}^{t+1}, \gamma_{2i+1}^{t+1}) \geq \left(\frac{u_{2i-1}^{t+1} + u_{2i+1}^{t+1}}{v_{2i-1}^{t+1} + v_{2i+1}^{t+1}} \right)^{a_{2i} + a_{2i+1}} h_{v_{2i}}(y | \mathcal{F}_t, \gamma_{2i-1}^{t+1}, \gamma_{2i+1}^{t+1})$$

hence

$$\begin{aligned} & \min \left\{ h_{u_{2i}} \left(y \mid \mathcal{F}_t, \gamma_{2i-1}^{t+1}, \gamma_{2i+1}^{t+1} \right), h_{v_{2i}} \left(y \mid \mathcal{F}_t, \gamma_{2i-1}^{t+1}, \gamma_{2i+1}^{t+1} \right) \right\} \\ & \geq \left(\frac{u_{2i-1}^{t+1} + u_{2i+1}^{t+1}}{v_{2i-1}^{t+1} + v_{2i+1}^{t+1}} \right)^{a_{2i} + a_{2i+1}} h_{v_{2i}} \left(y \mid \mathcal{F}_t, \gamma_{2i-1}^{t+1}, \gamma_{2i+1}^{t+1} \right) \end{aligned}$$

As described in the lead-up to Corollary 4.6, we take u_{2i}^{t+1} to be the x-coordinate of a uniformly chosen point from the area under the graph of $h_{u_{2i}} \left(y \mid \mathcal{F}_t, \gamma_{2i-1}^{t+1}, \gamma_{2i+1}^{t+1} \right)$, and set $v_{2i}^{t+1} = u_{2i}^{t+1}$ if this point is also below the graph $h_{v_{2i}} \left(y \mid \mathcal{F}_t, \gamma_{2i-1}^{t+1}, \gamma_{2i+1}^{t+1} \right)$. Otherwise take v_{2i}^{t+1} to be the x-coordinate of a uniformly chosen point from the area below the graph of $h_{v_{2i}} \left(y \mid \mathcal{F}_t, \gamma_{2i-1}^{t+1}, \gamma_{2i+1}^{t+1} \right)$ and above the graph of $h_{u_{2i}} \left(y \mid \mathcal{F}_t, \gamma_{2i-1}^{t+1}, \gamma_{2i+1}^{t+1} \right)$. The result is that

$$\begin{aligned} & \mathbb{P} \left[u_{2i}^{t+1} \neq v_{2i}^{t+1} \mid \mathcal{F}_t, \gamma_{2i-1}^{t+1}, \gamma_{2i+1}^{t+1} \right] \\ & = 1 - \int \min \left\{ h_{u_{2i}} \left(y \mid \mathcal{F}_t, \gamma_{2i-1}^{t+1}, \gamma_{2i+1}^{t+1} \right), h_{v_{2i}} \left(y \mid \mathcal{F}_t, \gamma_{2i-1}^{t+1}, \gamma_{2i+1}^{t+1} \right) \right\} dy \\ & \leq 1 - \left(\frac{u_{2i-1}^{t+1} + u_{2i+1}^{t+1}}{v_{2i-1}^{t+1} + v_{2i+1}^{t+1}} \right)^{a_{2i} + a_{2i+1}} \\ & \leq 1 - R_{t+1}^{-a_{2i} - a_{2i+1}} \\ & \leq 1 - R_t^{-a_{2i} - a_{2i+1}} \end{aligned}$$

where R_{t+1} is the process defined above and derived under the hypothetical continuation of the uniform coupling. Since the last inequality is independent of $\gamma_{2i-1}^{t+1}, \gamma_{2i+1}^{t+1}$,

we also get $\mathbb{P} [u_{2i}^{t+1} \neq v_{2i}^{t+1} | \mathcal{F}_t] \leq 1 - R_t^{-a_{2i}-a_{2i+1}}$. Therefore

$$\begin{aligned}
& \mathbb{P} \left[u^{t+1} \neq v^{t+1} \left| \mathcal{F}_t, \gamma_1^{t+1}, \gamma_3^{t+1}, \dots, \gamma_{2^{\lceil \frac{n}{2} \rceil - 1}}^{t+1} \right. \right] \\
&= \mathbb{P} \left[\cup_i \{u_{2i}^{t+1} \neq v_{2i}^{t+1}\} \left| \mathcal{F}_t, \gamma_1^{t+1}, \gamma_3^{t+1}, \dots, \gamma_{2^{\lceil \frac{n}{2} \rceil - 1}}^{t+1} \right. \right] \\
&= 1 - \prod_i \mathbb{P} \left[\{u_{2i}^{t+1} = v_{2i}^{t+1}\} \left| \mathcal{F}_t, \gamma_1^{t+1}, \gamma_3^{t+1}, \dots, \gamma_{2^{\lceil \frac{n}{2} \rceil - 1}}^{t+1} \right. \right] \\
&\leq 1 - \prod_i R_t^{-a_{2i}-a_{2i+1}} \\
&= 1 - R_t^{-\sum_i (a_{2i}+a_{2i+1})}
\end{aligned}$$

and once again since $R_t^{-\sum_i (a_{2i}+a_{2i+1})}$ is independent of $\left(\gamma_1^{t+1}, \gamma_3^{t+1}, \dots, \gamma_{2^{\lceil \frac{n}{2} \rceil - 1}}^{t+1} \right)$, we get the desired result. \square

4.5 Special case $n = 4$

In this section we will consider the Markov chain defined by the random functions $\{f_n^t\}$ in (4.4) for the case where $n = 4$, and with initial conditions $u^0 \preceq v^0$ that have the additional property $\frac{u_2^0}{u_4^0} = \frac{v_2^0}{v_4^0}$. Note from (4.7) that this may already be assumed. We proceed by defining the Markov chain u^t , and from here on we let v^t be the process defined in the following manner: set

$$(u_2^{t+1}, u_4^{t+1}) = f_4^{t+1}(u_2^t, u_4^t) = \left(f_{4,1}^{t+1}(u_2^t, u_4^t), f_{4,2}^{t+1}(u_2^t, u_4^t) \right)$$

and let

$$(\tilde{v}_2^{t+1}, \tilde{v}_4^{t+1}) = f_4^{t+1}(v_2^t, v_4^t) = \left(f_{4,1}^{t+1}(v_2^t, v_4^t), f_{4,2}^{t+1}(v_2^t, v_4^t) \right)$$

Then define

$$(v_2^{t+1}, v_4^{t+1}) := \max \left\{ \frac{\tilde{v}_2^{t+1}}{u_2^{t+1}}, \frac{\tilde{v}_4^{t+1}}{u_4^{t+1}} \right\} (u_2^{t+1}, u_4^{t+1}) \quad (4.10)$$

Observe that:

1. The equality of the ratios is always preserved: $\frac{u_2^{t+1}}{u_4^{t+1}} = \frac{v_2^{t+1}}{v_4^{t+1}}$.
2. $(\tilde{v}_2^{t+1}, \tilde{v}_4^{t+1}) \preceq (v_2^{t+1}, v_4^{t+1})$, hence by monotonicity the process v^t is always

“greater than or equal to” a copy of the Markov chain started at v^0 and coupled uniformly with u^t .

3. From the ratio above we also get $\frac{v_2^{t+1}}{u_2^{t+1}} = \frac{v_4^{t+1}}{u_4^{t+1}}$.

Hence if we let $R_t := \frac{v_2^t}{u_2^t} = \frac{v_4^t}{u_4^t}$, then

$$\begin{aligned} \frac{v_2^{t+1}}{u_2^{t+1}} &= \max \left\{ \frac{\tilde{v}_2^{t+1}}{u_2^{t+1}}, \frac{\tilde{v}_4^{t+1}}{u_4^{t+1}} \right\} \\ &= \max \left\{ \left(\frac{\frac{\gamma_3^{t+1}}{u_2^t + u_4^t} + b}{\frac{\gamma_3^{t+1}}{v_2^t + v_4^t} + b} \right), \left(\frac{\frac{\gamma_3^{t+1}}{u_2^t + u_4^t} + \frac{\gamma_1^{t+1}}{u_2^t + x}}{\frac{\gamma_3^{t+1}}{v_2^t + v_4^t} + \frac{\gamma_1^{t+1}}{v_2^t + x}} \right) \right\} \\ &= \frac{v_2^t}{u_2^t} \cdot \max \left\{ \left(\frac{\frac{\gamma_3^{t+1}}{\frac{u_2^t}{u_4^t} + 1} + b u_4^t}{\frac{\gamma_3^{t+1}}{\frac{u_2^t}{u_4^t} + 1} + b v_4^t} \right), \left(\frac{\frac{\gamma_3^{t+1}}{\frac{u_4^t}{u_2^t} + 1} + \frac{\gamma_1^{t+1}}{1 + \frac{x}{u_2^t}}}{\frac{\gamma_3^{t+1}}{\frac{u_4^t}{u_2^t} + 1} + \frac{\gamma_1^{t+1}}{1 + \frac{x}{v_2^t}}} \right) \right\} \\ &\leq \frac{v_2^t}{u_2^t} Q_t \end{aligned} \quad (4.11)$$

where

$$Q_t = \max \left\{ \left(\frac{\gamma_3^{t+1} + \frac{\gamma_1^{t+1}}{1 + \frac{x}{u_2^t}}}{\gamma_3^{t+1} + \frac{\gamma_1^{t+1}}{1 + \frac{x}{v_2^t}}} \right), \left(\frac{\gamma_3^{t+1} + b u_4^t}{\gamma_3^{t+1} + b v_4^t} \right) \right\}$$

$$\leq 1$$

The inequality in (4.11) is justified by

Lemma 4.8. *Suppose that $0 < a < b$. Then $g(\mathbf{x}, \mathbf{y}) := (\frac{\mathbf{x}}{b} + \mathbf{y}) / (\frac{\mathbf{x}}{a} + \mathbf{y})$ is decreasing in \mathbf{x} and increasing in \mathbf{y} , for all $\mathbf{x}, \mathbf{y} > 0$.*

Proof. Follows from calculus.

□

Therefore

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

It can be easily observed that the ratio R_t satisfies the condition stated in the paragraph preceding Lemma 4.7, namely that $R_t \geq \max \left\{ \frac{v_i^t}{u_i^t} \right\}$. The aim now is to obtain from the previous inequality 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 (4.21)). Note that in order to achieve this, it suffices to have for all $t \geq 0$

$$\mathbb{E}[Q_t R_t] \leq r_{t+1} (\mathbb{E}[R_t] - 1) + 1 \tag{4.12}$$

Let $\mathcal{F}_t := \sigma((\gamma_1^1, \dots, \gamma_4^1), \dots, (\gamma_1^t, \dots, \gamma_4^t))$. We can consider (4.12) by conditioning on this filtration

$$\mathbb{E}[Q_t R_t] = \mathbb{E} \left[R_t \mathbb{E} [Q_t | \mathcal{F}_t] \right] \quad (4.13)$$

and we may approximate $\mathbb{E} [Q_t | \mathcal{F}_t]$ with the aid of the following lemmas.

Let $\mu_1 = \mathbb{E} [\gamma_3] = a_3 + a_4$ and $\mu_2 = \mathbb{E} [\gamma_1 - \frac{1}{3}] = a_1 + a_2 - \frac{1}{3}$.

Lemma 4.9. *Let S be a \mathcal{F}_t -measurable stopping time. Then*

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

where $\dot{r}_t = 1 - 1/\max \left\{ \left(\frac{4\mu_1}{\mu_2} + 4 \right) \left(\frac{u_2^t}{x} + \frac{x}{v_2^t} + 2 \right), 4 + \frac{4\mu_1}{bv_4^t} \right\}$.

Proof. By [4] 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 with probability of at least $\frac{1}{4}$ we have that $Q_t \leq \max \left\{ \left(\frac{\mu_1 + \frac{\mu_2}{1 + \frac{x}{v_2^t}}}{\mu_1 + \frac{\mu_2}{1 + \frac{x}{v_2^t}}} \right), \left(\frac{\mu_1 + bv_4^t}{\mu_1 + bv_4^t} \right) \right\}$ for any t . Then by

the previous lemma

$$\begin{aligned}
& \mathbb{E} [Q_S \mid \mathcal{F}_S] \\
& \leq \frac{1}{4} \cdot \max \left\{ \left(\frac{\mu_1 + \frac{\mu_2}{1 + \frac{x}{u_2^S}}}{\mu_1 + \frac{\mu_2}{1 + \frac{x}{v_2^S}}} \right), \left(\frac{\mu_1 + bu_4^S}{\mu_1 + bv_4^S} \right) \right\} + 1 \cdot \frac{3}{4} \\
& = \frac{1}{4} \cdot \max \left\{ \left(1 - \frac{\frac{1}{1 + \frac{x}{v_2^S}} - \frac{1}{1 + \frac{x}{u_2^S}}}{\frac{\mu_1}{\mu_2} + \frac{1}{1 + \frac{x}{v_2^S}}} \right), \left(1 - \frac{bv_4^S - bu_4^S}{\mu_1 + bv_4^S} \right) \right\} + 1 \cdot \frac{3}{4} \\
& \leq \frac{1}{4} \cdot \max \left\{ 1 - \frac{\left(1 - \frac{1}{R_S}\right)}{\left(\frac{\mu_1}{\mu_2} + 1\right) \left(1 + \frac{x}{v_2^S}\right) \left(1 + \frac{x}{u_2^S}\right) \frac{u_2^S}{x}}, 1 - \frac{\left(1 - \frac{1}{R_S}\right)}{1 + \frac{\mu_1}{bv_4^S}} \right\} + 1 \cdot \frac{3}{4} \\
& \leq 1 - \frac{\left(1 - \frac{1}{R_S}\right)}{\max \left\{ \left(\frac{4\mu_1}{\mu_2} + 4\right) \left(\frac{u_2^S}{x} + \frac{x}{v_2^S} + 2\right), 4 + \frac{4\mu_1}{bv_4^S} \right\}} \\
& = \dot{r}_S + \frac{1 - \dot{r}_S}{R_S} \tag{4.14}
\end{aligned}$$

Substituting this into (4.13), we get the desired result. \square

The next step will be to show that often enough $\dot{r}_t \leq r$ for some $r < 1$, which by the inequality (4.14) would result in an expression of the form given by (4.12).

4.6 Super-martingale-type auxiliary processes

Most results in this section are not restricted to only the Markov chain in (4.4), but can be applied more generally. In particular, we will make use of them again for the case $n > 4$ with an alternatively defined Markov chain. Lemma 4.10 holds

for any adapted processes $K_{i,t}$ satisfying the inequality (4.15), despite the fact that here it is considered as a function of (u^t, v^t) . Lemma 4.11 is a general observation, independent of anything discussed so far. Lemma 4.12 relies on the same premise as Lemma 4.10, but here we do make use of the fact that $K_{i,t}$ is a function of (u^t, v^t) . The definition and dimension of the random variable (u^t, v^t) is irrelevant in the proof, and we only require that it satisfies the Markov property. Lemma 4.13 makes explicit reference to the process D_t which is defined in the beginning of Section (4.7) , however the result remains true for any adapted process which satisfies the necessary conditions given in (4.17) and (4.18), and is used to bound R_t . Lastly, the transition from Lemma 4.13 to Corollary 4.14 is only based on some simplification of binomial coefficients, hence it will also translate easily to the case $n > 4$.

To illustrate the relevance of these results to the problem we have built so far (i.e. for $n = 4$), we will start by assuming the existence of a set of auxiliary processes which satisfy conditions outlined below, and which can serve to provide an upper bound to the random part of \dot{r}_t , namely to

$\max \left\{ \left(\frac{4\mu_1}{\mu_2} + 4 \right) \left(\frac{u_2^t}{x} + \frac{x}{v_2^t} + 2 \right), 4 + \frac{4\mu_1}{bv_4^t} \right\}$. Suppose that for $i = 1, \dots, N$, the processes $K_{i,t} = K_i(u^t, v^t)$ are adapted to \mathcal{F}_t , and that for $t \geq 0$

$$\mathbb{E} [K_{i,t+1} | \mathcal{F}_t] \leq \zeta_i K_{i,t} + c_i \tag{4.15}$$

where $\zeta_i < 1$ and c_i are constants. Then for the process $J_t = J(u^t, v^t) := \sum_i K_{i,t}$

we also have

$$\mathbb{E} [J_{t+1} | \mathcal{F}_s] \leq \max \{\zeta_i\} J_t + \sum_i c_i$$

In particular, if $J_s \geq \eta := \frac{2 \sum c_i}{1 - \max \{\zeta_i\}}$ for some $s \geq 0$, then $\mathbb{E} [J_{s+1} | \mathcal{F}_s] \leq \beta J_s$ where

$\beta := \frac{1 + \max \{\zeta_i\}}{2}$. Let

$T = T(s) := \min \{\tau > s | J_\tau \leq \eta\}$, and for $s \geq 0$ and $t \geq 1$ define

$$\hat{J}_{s,t} := \begin{cases} J_{s+t} & s+t < T, J_s \leq \eta \\ 0 & \text{otherwise} \end{cases}$$

or in other words $\hat{J}_{s,t} = \mathbf{1}_{\{J_s \leq \eta\} \cap \{T > s+t\}} J_{s+t}$.

Lemma 4.10. *For the notation and assumptions of the preceding paragraph,*

$$\mathbb{E} [\hat{J}_{s,t+1} | \mathcal{F}_s] \leq \beta^{t+1} \eta \text{ for } t \geq 0 \text{ and } s \geq 0.$$

Proof. Observe that for $t \geq 1$,

$$\begin{aligned} \mathbb{E} [\hat{J}_{s,t+1} | \mathcal{F}_{s+t}] &= \mathbf{1}_{\{J_s \leq \eta\} \cap \{T \leq s+t\}} \mathbb{E} [\hat{J}_{s,t+1} | \mathcal{F}_{s+t}] + \mathbf{1}_{\{J_s \leq \eta\} \cap \{T > s+t\}} \mathbb{E} [\hat{J}_{s,t+1} | \mathcal{F}_{s+t}] \\ &= 0 + \mathbf{1}_{\{J_s \leq \eta\} \cap \{T > s+t\}} \mathbb{E} [\mathbf{1}_{\{T > s+t+1\}} J_{s+t+1} | \mathcal{F}_{s+t}] \\ &\leq \mathbf{1}_{\{J_s \leq \eta\} \cap \{T > s+t\}} \mathbb{E} [\mathbf{1}_{\{J_{s+t} > \eta\}} J_{s+t+1} | \mathcal{F}_{s+t}] \\ &\leq \mathbf{1}_{\{J_s \leq \eta\} \cap \{T > s+t\}} \beta J_{s+t} \\ &= \beta \hat{J}_{s,t} \end{aligned}$$

Proceeding inductively, it follows that

$$\mathbb{E} [\hat{J}_{s,t+1} | \mathcal{F}_s] \leq \mathbb{E} [\beta^t \hat{J}_{s,1} | \mathcal{F}_s]$$

Finally,

$$\begin{aligned}
\mathbb{E} \left[\hat{J}_{s,t} \mid \mathcal{F}_s \right] &\leq \mathbb{E} \left[\mathbf{1}_{\{J_s \leq \eta\}} J_{s+1} \mid \mathcal{F}_s \right] \\
&= \mathbb{E} \left[J_{s+1} \mid \mathcal{F}_s \right] \mathbf{1}_{\{J_s \leq \eta\}} \\
&\leq \left(\sum_i c_i + \max \{ \zeta_i \} J_s \right) \mathbf{1}_{\{J_s \leq \eta\}} \\
&\leq \left(\sum_i c_i + \max \{ \zeta_i \} \eta \right) \mathbf{1}_{\{J_s \leq \eta\}} \\
&\leq \beta \eta
\end{aligned}$$

□

Remark 4.1. *If it is uncertain that $J_s \leq \eta$, we can still define $\check{J}_{s,t} = \mathbf{1}_{\{T > s+t\}} J_{s+t}$, and following the proof of Lemma 4.10 it is a straight forward conclusion that*

$$\mathbb{E} \left[\check{J}_{s,t+1} \mid \mathcal{F}_s \right] \leq \beta^{t+1} \max \{ \eta, J_s \} \tag{4.16}$$

Now suppose that D_t is a process adapted to \mathcal{F}_t such that

$$\forall t \geq 1, D_t \geq \max \left\{ \left(\frac{4\mu_1}{\mu_2} + 4 \right) \left(\frac{u_2^t}{x} + \frac{x}{v_2^t} + 2 \right), 4 + \frac{4\mu_1}{bv_4^t} \right\} \tag{4.17}$$

Furthermore, suppose also that

$$D_{t+1} \leq \omega_{N+1,t+1} + \sum_{i=1}^N \omega_{i,t+1} K_{i,t} \tag{4.18}$$

where $(\omega_{1,t+1}, \dots, \omega_{N+1,t+1})$ is a non-negative random vector, i.i.d. over time $t \geq 1$, measurable w.r.t. \mathcal{F}_{t+1} and independent of \mathcal{F}_t . It is now clear that D_t is defined

with the intent to serve as an upper bound for \dot{r}_t . We will construct D_t in Section 4.7, and the reasons for insisting on the condition given in (4.18) will become apparent.

If S is a finite a.s. stopping time adapted to \mathcal{F}_t s.t. $J_S \leq \eta$, then $D_{S+1} \leq \eta \sum \omega_{i,S+1} + \omega_{N+1,S+1}$. Therefore, applying Lemma 4.9 we get

$$\begin{aligned}
\mathbb{E}[R_{S+2}] &\leq \mathbb{E}[Q_{S+1}R_{S+1}] \\
&\leq \mathbb{E}[\dot{r}_{S+1}(R_{S+1} - 1)] + 1 \\
&\leq \mathbb{E}\left[\left(1 - \frac{1}{D_{S+1}}\right)(R_{S+1} - 1)\right] + 1 \\
&\leq \mathbb{E}\left[\left(1 - \frac{1}{D_{S+1}}\right)(R_S - 1)\right] + 1 \\
&\leq \mathbb{E}\left[\left(1 - \frac{1}{\eta \sum \omega_{i,S+1} + \omega_{N+1,S+1}}\right)(R_S - 1)\right] + 1 \\
&= \mathbb{E}\left[\left(1 - \frac{1}{\eta \sum \omega_{i,S+1} + \omega_{N+1,S+1}}\right)\right] \mathbb{E}[(R_S - 1)] + 1 \\
&\leq r \mathbb{E}[(R_S - 1)] + 1
\end{aligned} \tag{4.19}$$

where $r = 1 - 1/((\theta_1 + \dots + \theta_N)\eta + \theta_{N+1})$ and $\theta_i := \mathbb{E}[\omega_{i,t+1}]$. Here we have used Jensen's inequality in the transition between the last two lines in (4.19). An additional observation that will be useful to us later, is that if $0 \leq Y \in \mathcal{F}_s$ then by a derivation identical to (4.19) we get

$$\mathbb{E}[Y R_{S+2}] \leq r \mathbb{E}[Y (R_S - 1)] + \mathbb{E}[Y] \tag{4.20}$$

The term $\mathbb{E}[Y]$ in the right-hand side of (4.20) comes about in the second line of (4.19), as a result of applying (4.14).

Lemma 4.11. *Let Y be a random variable. If A is an event and $B \subseteq \mathbb{R}$,*

$$\mathbb{P}[A|Y \in B] \leq \sup_{y_0 \in B} \mathbb{P}[A|Y = y_0]$$

Proof. Suppose on the contrary that $\mathbb{P}[A|Y \in B] > \mathbb{P}[A|Y = y_0]$ for all y_0 in B . Multiplying both sides by the marginal density $f_Y(y_0)$ and integrating over y_0 in B gives the contradiction

$$\mathbb{P}[A, Y \in B] > \mathbb{P}[A, Y \in B]$$

□

Working under the assumption that a process D_t satisfying the aforementioned conditions exists, and that Lemma 4.10 applies, we define the set \bar{S}_t by

$$\bar{S}_t := \{1 \leq i \leq t | J_i \leq \eta\} \tag{4.21}$$

Then

Lemma 4.12. *For any subset $\{c_1, c_2, \dots, c_k\} \subseteq \{1, \dots, t\}$,*

$$\mathbb{P}[\bar{S}_t = \{c_1, c_2, \dots, c_k\} | J_0 \leq \eta] \leq \beta^{t-k}$$

For the purpose of this lemma we will consider the previously defined function J as a function on \mathbb{R}_+^8 , as this will allow us to refer to the odd numbered co-ordinates u_1^t, u_3^t, v_1^t and v_3^t when we later define a satisfactory auxiliary process. It will be clear that this interpretation has no impact on any of previously derived results (such as Lemma 4.10) that we may need to refer to.

Proof. Let $A = \{(y_1, \dots, y_8) \in \mathbb{R}_+^8 \text{ s.t. } J(y_1, \dots, y_8) \leq \eta\}$, and $I \subseteq \{0, 1, \dots, k\}$ be those indices i that satisfy $c_{i+1} > c_i + 1$, where by convention we set $c_0 = 0$ and $c_{k+1} = t + 1$. Then for $i \in I$ let $B_i = \{J_{c_{i+1}} > \eta, \dots, J_{c_{i+1}-1} > \eta\}$. By Lemma 4.11,

$$\mathbb{P} \left[B_i \mid J_{c_i} \leq \eta \right] \leq \sup_{y \in A} \mathbb{P} \left[B_i \mid (u^{c_i}, v^{c_i}) = y \right]$$

and since J_{c_i} is determined by the values (u^{c_i}, v^{c_i}) , it follows by the same reasoning and the Markov property that also for any event $C_{c_{i-1}} \in \mathcal{F}_{c_{i-1}}$

$$\mathbb{P} \left[B_i \mid J_{c_i} \leq \eta, C_{c_{i-1}} \right] \leq \sup_{y \in A} \mathbb{P} \left[B_i \mid (u^{c_i}, v^{c_i}) = y \right] \quad (4.22)$$

Observe also that if $I = \{i_1, \dots, i_m\}$ for some $m \leq k + 1$, then

$$\begin{aligned} \sum_{j=1}^m (c_{i_j+1} - c_{i_j} - 1) &= |\{1, \dots, t\} \setminus \{c_1, c_2, \dots, c_k\}| \\ &= t - k \end{aligned}$$

Hence we get

$$\begin{aligned}
& \mathbb{P} [\bar{S}_t = \{c_1, c_2, \dots, c_k\} | J_0 \leq \eta] \\
&= \mathbb{P} \left[\{J_{c_1} \leq \eta, \dots, J_{c_k} \leq \eta\} \cap \{\cap_{i \in I} B_i\} | J_0 \leq \eta \right] \\
&\leq \mathbb{P} \left[\{J_{c_{i_1}} \leq \eta, \dots, J_{c_{i_m}} \leq \eta\} \cap \{\cap_{i \in I} B_i\} | J_0 \leq \eta \right] \\
&= \mathbb{P} \left[B_{i_m} \left| \{J_{c_{i_1}} \leq \eta, \dots, J_{c_{i_m}} \leq \eta\} \cap \left\{ \cap_{j=1}^{m-1} B_{i_j} \right\}, J_0 \leq \eta \right. \right] \cdot \\
&\quad \mathbb{P} \left[\{J_{c_{i_1}} \leq \eta, \dots, J_{c_{i_m}} \leq \eta\} \cap \left\{ \cap_{j=1}^{m-1} B_{i_j} \right\} | J_0 \leq \eta \right] \\
&\leq \sup_{y \in A} \mathbb{P} \left[B_{i_m} | (u^{c_{i_m}}, v^{c_{i_m}}) = y \right] \cdot \mathbb{P} \left[\{J_{c_{i_1}} \leq \eta, \dots, J_{c_{i_{m-1}}} \leq \eta\} \cap \left\{ \cap_{j=1}^{m-1} B_{i_j} \right\} | J_0 \leq \eta \right] \\
&\quad \vdots \\
&\leq \prod_{j=1}^m \sup_{y \in A} \mathbb{P} \left[B_{i_j} | (u^{c_{i_j}}, v^{c_{i_j}}) = y \right] \\
&\leq \prod_{j=1}^m \sup_{y \in A} \mathbb{P} \left[\hat{J}_{c_{i_j}, c_{(i_j+1)} - c_{i_j} - 1} > \eta | (u^{c_{i_j}}, v^{c_{i_j}}) = y \right] \\
&\leq \beta^{c_{i_1+1} - c_{i_1} - 1} \dots \beta^{c_{i_m+1} - c_{i_m} - 1} \\
&= \beta^{t-k}
\end{aligned}$$

The inequality before the last line follows from Lemma 4.10 and Markov's inequality.

We note that when $i_1 = 0$, the event $J_{i_1} \leq \eta$ appears in the second line, but not in the first. This can be justified by observing that in this case $J_{i_1} = J_0$, and $J_0 \leq \eta$ is already given in this conditional probability. \square

From Lemma 4.12 it also immediately follows that

$$\mathbb{P} \left[|\bar{S}_t| = k \mid J_0 \leq \eta \right] \leq \binom{t}{k} \beta^{t-k} \quad (4.23)$$

Lemma 4.13. *Assuming that the process D_t is adapted to \mathcal{F}_t and satisfies (4.17) and (4.18). Then in the event $\{J_0 \leq \eta\}$*

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

Proof. Let $\tau_0 = 0$ and $\{\tau_i\} \subseteq \{1, 2, \dots\}$ be those times for which $J_{\tau_i} \leq \eta$. Then by

(4.20) with $Y = \mathbf{1}_{\tau_{k+i} \leq t}$ and $S = \tau_{k+1}$

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

The last inequality uses the fact that $\mathbf{1}_{\tau_{k+1} \leq t} \leq \mathbf{1}_{\tau_{k-1} \leq t}$ and $R_{\tau_{k+1}} \leq R_{\tau_{k-1}+2}$. This

then leads to the first step in an inductive argument:

$$\begin{aligned} &\mathbb{E} \left[R_{\tau_{k+1}+2} \mathbf{1}_{\tau_{k+1} \leq t} \mid \mathcal{F}_0 \right] - \mathbb{P} \left[|\bar{S}_t| > k \mid \mathcal{F}_0 \right] \\ &\leq r \left(\mathbb{E} \left[R_{\tau_{k-1}+2} \mathbf{1}_{\tau_{k-1} \leq t} \mid \mathcal{F}_0 \right] - \mathbb{P} \left[|\bar{S}_t| > k-2 \mid \mathcal{F}_0 \right] \right) \end{aligned} \quad (4.24)$$

Proceeding in this manner, we claim that we get

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

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, by (??)

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

The second line follows from (4.20). □

From (4.23) and Lemma 4.13 we conclude

$$\begin{aligned} \mathbb{E} [R_{t+2} | \mathcal{F}_0] &= \mathbb{E} [R_{t+2} \mathbf{1}_{|\bar{S}_t| > k} | \mathcal{F}_0] + \mathbb{E} [R_{t+2} \mathbf{1}_{|\bar{S}_t| \leq k} | \mathcal{F}_0] \\ &\leq r^{\lceil (k+1)/2 \rceil} (R_0 - 1) + \mathbb{P} [|\bar{S}_t| > k | \mathcal{F}_0] + \mathbb{E} [R_0 \mathbf{1}_{|\bar{S}_t| \leq k} | \mathcal{F}_0] \\ &\leq r^{\lceil (k+1)/2 \rceil} (R_0 - 1) + \mathbb{P} [|\bar{S}_t| > k | \mathcal{F}_0] + \\ &\quad + (R_0 - 1) \mathbb{P} [|\bar{S}_t| \leq k | \mathcal{F}_0] + \mathbb{P} [|\bar{S}_t| \leq k | \mathcal{F}_0] \\ &\leq 1 + (R_0 - 1) \left(r^{\lceil (k+1)/2 \rceil} + \sum_{j=0}^k \binom{t}{j} \beta^{t-j} \right) \end{aligned} \tag{4.25}$$

Inequality (4.25) is true for any $k \leq t$, so we are free to choose any value for k in this range. We can simplify this expression by removing the binomial terms in the following manner: note first that if we take $k \leq \lfloor \frac{t}{3} \rfloor$ then for $j < k$, $\binom{t}{j} \leq \frac{1}{2} \binom{t}{j+1}$, hence

$\sum_{j=0}^k \binom{t}{j} \beta^{t-j} \leq 2 \binom{t}{k} \beta^{t-k}$. Next, let $d := \lceil \frac{t}{k} \rceil$ and note that $\binom{dk}{k} q^k (1-q)^{dk-k} \leq 1$ whenever $q \in (0, 1)$. Therefore, if $d \geq 2$ then by taking $q = \frac{1}{d}$ we get

$$\binom{t}{k} \leq \binom{dk}{k} \leq \frac{d^{dk}}{(d-1)^{(d-1)k}} \quad (4.26)$$

From these remarks and conditions, it follows that the summation in (4.25) may be replaced by

$2 \left(\frac{d^d}{(d-1)^{(d-1)}} \right)^k \beta^{t-k}$. Our goal is to bound d from below by a constant d' (and thereby set k to be a fraction of t) in such a way that $\left(\frac{d^d}{(d-1)^{(d-1)}} \right)^k \beta^{t-k}$ is decaying exponentially in t . Since $\left(\frac{d^d}{(d-1)^{(d-1)}} \right)^k \beta^{t-k} = \left(\left(\frac{d^d}{(d-1)^{(d-1)}} \right)^{t/k-1} \right)^k$, this aim would be achieved if we could find d' such that for $d \geq d'$

$$\frac{d^d}{(d-1)^{(d-1)}} \beta^{(d-2)} \leq \sqrt{r} \quad (4.27)$$

Here the term \sqrt{r} is chosen for convenience, and (4.27) would then imply that

$$\sum_{j=0}^k \binom{t}{j} \beta^{t-j} \leq 2 \left(\left(\frac{d^d}{(d-1)^{(d-1)}} \right)^{t/k-1} \right)^k \leq 2r^{k/2}$$

The left hand side of (4.27) is equal to $d \left(1 + \frac{1}{d-1} \right)^{d-1} \beta^{(d-2)} \leq ed\beta^{(d-2)}$. Hence

(4.27) is true if

$d \geq \frac{\ln\left(\frac{\sqrt{r}}{ed}\right)}{\ln(\beta)} + 2 = \frac{1}{|\ln(\beta)|} \left\{ \ln(d) + \ln\left(\frac{e}{\sqrt{r}}\right) \right\} + 2$. Since $\ln(d) \leq \sqrt{d}$, we can consider the inequality $d \geq \frac{1}{|\ln(\beta)|} \sqrt{d} + \frac{\ln\left(\frac{e}{\sqrt{r}}\right)}{|\ln(\beta)|} + 2$ or $\sqrt{d} \geq \frac{1}{|\ln(\beta)|} + \frac{\ln\left(\frac{e}{\sqrt{r}}\right)}{|\ln(\beta)|\sqrt{d}} + \frac{2}{\sqrt{d}}$, which is certainly true if $d \geq \left(\frac{1}{|\ln(\beta)|} + \frac{1}{|\ln(\beta)|} \ln\left(\frac{e}{\sqrt{r}}\right) + 2 \right)^2$. Therefore taking $d' :=$

$\left(\frac{1}{|\ln(\beta)|} + \frac{1}{|\ln(\beta)|} \ln\left(\frac{e}{\sqrt{r}}\right) + 2\right)^2$ and setting $k = \lfloor \frac{t}{d'} \rfloor$ (note that the condition $k \leq \lfloor \frac{t}{3} \rfloor$ from the previous paragraph is satisfied), we get that $d = \lceil \frac{t}{k} \rceil \geq d'$, hence

$$\begin{aligned} 1 + (R_0 - 1) \left(r^{\lceil (k+1)/2 \rceil} + \sum_{j=0}^k \binom{t}{j} \beta^{t-j} \right) &\leq 1 + (R_0 - 1) \left(r^{\lceil (k+1)/2 \rceil} + 2r^{k/2} \right) \\ &\leq 1 + r^{\frac{1}{2} \lfloor \frac{t}{d'} \rfloor} (R_0 - 1) (\sqrt{r} + 2) \end{aligned} \quad (4.28)$$

We summarise this in the following corollary

Corollary 4.14. $\mathbb{E} [R_{t+2} | \mathcal{F}_0] \leq 1 + 3r^{\frac{1}{2} \lfloor \frac{t}{d'} \rfloor} (R_0 - 1)$ for $t \geq d'$, where d' and r are as defined previously.

4.7 Construction of D_t

For ease of reference, we will start by first giving the following list of definitions

$K_{1,t} := u_2^t + u_4^t$	$K_{2,t} := \frac{u_3^t + u_1^t + b}{\gamma_2^t + \gamma_4^t}$
$C_2 := \frac{a_1 + a_2 + xb}{x(a_2 + a_3 + a_4 + a_5 - 1)}$	
$D_t := \frac{\zeta}{x} (u_2^t + u_4^t) + \left(\zeta x + \frac{4\mu_1}{b} \right) \left(\frac{1}{u_2^t} + \frac{1}{u_4^t} \right)$	
$\zeta_1 := \frac{a_2 + a_3}{a_1 + a_2 + a_3 + a_4 - 1}$	$\zeta_2 := \frac{a_3 + a_4}{a_2 + a_3 + a_4 + a_5 - 1}$
$C_1 := \zeta_1 x + \frac{a_4 + a_5}{b}$	$\varsigma := \frac{4\mu_1}{\mu_2} + 4$
$\tilde{\omega}_{2,t+1} = 2 + \frac{\gamma_2^{t+1}}{\gamma_4^{t+1}} + \frac{\gamma_4^{t+1}}{\gamma_2^{t+1}}$	$\omega_{1,t+1} := \frac{\zeta}{x} \frac{\gamma_2^{t+1}}{\gamma_1^{t+1} + \gamma_3^{t+1}}$
$\omega_{2,t+1} := \left(\zeta x + \frac{4\mu_1}{b} \right) \tilde{\omega}_{2,t+1} \frac{\gamma_3^{t+1}}{\gamma_2^{t+1} + \gamma_4^{t+1}}$	
$\omega_{3,t+1} := \frac{\zeta}{x} \left(\frac{\gamma_2^{t+1}}{\gamma_1^{t+1} + \gamma_3^{t+1}} x + \frac{\gamma_4^{t+1}}{b} \right) + \left(\zeta x + \frac{4\mu_1}{b} \right) \tilde{\omega}_{2,t+1} \frac{\frac{\gamma_1^{t+1}}{x} + b}{\gamma_2^{t+1} + \gamma_4^{t+1}}$	

Note that

$$\max \left\{ \left(\frac{4\mu_1}{\mu_2} + 4 \right) \left(\frac{u_2^t}{x} + \frac{x}{v_2^t} + 2 \right), 4 + \frac{4\mu_1}{bv_4^t} \right\} \leq D_t \quad (4.29)$$

where we have used the fact that $u \preceq v$ and $2 \leq \frac{u}{x} + \frac{x}{u}$. To bound the first term in this sum, observe that

$$\begin{aligned} u_2^{t+1} + u_4^{t+1} &= \frac{\gamma_2^{t+1}}{\frac{\gamma_1^{t+1}}{x+u_2^t} + \frac{\gamma_3^{t+1}}{u_2^t+u_4^t}} + \frac{\gamma_4^{t+1}}{\frac{\gamma_3^{t+1}}{u_2^t+u_4^t} + b} \\ &\leq \frac{\gamma_2^{t+1}}{\gamma_1^{t+1} + \gamma_3^{t+1}} (u_2^t + u_4^t + x) + \frac{\gamma_4^{t+1}}{b} \end{aligned} \quad (4.30)$$

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

$$\begin{aligned} u_3^{t+1} &= \frac{\gamma_3^{t+1}}{u_2^t + u_4^t} = \frac{\gamma_3^{t+1}}{\frac{\gamma_2^t}{u_1^t+u_3^t} + \frac{\gamma_4^t}{u_3^t+b}} \\ &\leq \frac{\gamma_3^{t+1}}{\gamma_2^t + \gamma_4^t} (u_1^t + u_3^t + b) \end{aligned}$$

it follows that

$$K_{2,t+1} \leq \frac{\gamma_3^{t+1}}{\gamma_2^{t+1} + \gamma_4^{t+1}} K_{2,t} + \frac{u_1^{t+1} + b}{\gamma_2^{t+1} + \gamma_4^{t+1}} \quad (4.31)$$

and hence

$$\begin{aligned} \mathbb{E} [K_{2,t+1} | \mathcal{F}_t] &\leq \zeta_2 K_{2,t} + \mathbb{E} \left[\frac{\frac{\gamma_3^{t+1}}{x} + b}{\gamma_2^{t+1} + \gamma_4^{t+1}} | \mathcal{F}_t \right] \\ &\leq \zeta_2 K_{2,t} + C_2 \end{aligned} \quad (4.32)$$

Both $K_{1,t}$ and $K_{2,t}$ are adapted to \mathcal{F}_t and are in fact functions of 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 $t \geq 1$, and we set $K_{2,0} = u_2^0 (u_1^0 + u_3^0) + u_4^0 (u_3^0 + b)$).

Note also that

$$\begin{aligned} \frac{1}{u_2^{t+1}} + \frac{1}{u_4^{t+1}} &\leq \left(\frac{1}{\gamma_2^{t+1}} + \frac{1}{\gamma_4^{t+1}} \right) (u_1^{t+1} + u_3^{t+1} + b) \\ &= \tilde{\omega}_{2,t+1} K_{2,t+1} \end{aligned} \quad (4.33)$$

and $\tilde{\omega}_{2,t+1}$ is independent of \mathcal{F}_t . By (4.30), (4.31) and (4.33) we conclude that

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

and hence D_t satisfies the conditions given by and preceding equation (4.18). Referring back to (4.19), we obtain the rate

$$r = 1 - \frac{1}{(\theta_1 + \theta_2)\eta + \theta_3} \quad (4.34)$$

where $\theta_1, \theta_2, \theta_3$ are the expected values of $\omega_{1,t+1}, \omega_{2,t+1}$ and $\omega_{3,t+1}$ respectively.

We make the additional note that it is not necessary for $\{K_{i,t}\}$ to be deterministic functions of (u^t, v^t) . This assumption was required to make use of the Markov property in (4.19) and (4.22), however the arguments remain true if $\{K_{i,t}\}$ are random functions of (u^t, v^t) with random terms that are independent of \mathcal{F}_∞ .

Note also that condition (4.6) guarantees that $\zeta_1 < 1$ and $\zeta_2 < 1$, as well as the finite value of all constants and finite expectation of all random variables defined in the beginning of this section.

We have now established a sufficient foundation to prove our first theorem.

Proof of Theorem 4.2. By Corollary 4.6, $\mathbb{P}[u^{t+3} \neq v^{t+3}]$ is an upper bound for $d_{TV}(\mathcal{U}^{t+3}, \mathcal{V}^{t+3})$ under the specified 'one shot' coupling described in the paragraph preceding the corollary (in the aforementioned description we couple uniformly until time t , and attempt to merge the two Markov chains thereafter. Here we attempt to do this after time $t + 2$, but the argument remains the same). In the event $\{J_0 \leq \eta\}$ we conclude by Corollary 4.14

$$\mathbb{E}[R_{t+2} - 1 | \mathcal{F}_0,] \leq 3r^{\frac{1}{2}} \lfloor \frac{t}{d'} \rfloor (R_0 - 1)$$

Therefore by Lemma 4.7 and using Jensen's inequality

$$\begin{aligned} \mathbb{P}[u^{t+3} \neq v^{t+3} | \mathcal{F}_0] &= \mathbb{E}\left[\mathbb{P}[u^{t+3} \neq v^{t+3} | \mathcal{F}_{t+2}] | \mathcal{F}_0, \right] \\ &\leq \mathbb{E}\left[1 - (R_{t+2})^{-(a_2+a_3+a_4+a_5)}\right] \\ &\leq 1 - (\mathbb{E}[R_{t+2}])^{-(a_2+a_3+a_4+a_5)} \\ &\leq 1 - \left(1 + 3r^{\frac{1}{2}} \lfloor \frac{t}{d'} \rfloor (R_0 - 1)\right)^{-(a_2+a_3+a_4+a_5)} \end{aligned} \quad (4.35)$$

We claim that the right-hand side of (4.35) is bounded by

$3r^{\frac{1}{2}} \lfloor \frac{t}{d'} \rfloor (a_2 + a_3 + a_4 + a_5) (R_0 - 1)$. To justify this claim, define $E(y) := \frac{1}{(1+y)^\nu} + \nu y$ for $y, \nu \in \mathbb{R}^+$, and observe that $E'(y) = \frac{-\nu}{(1+y)^{\nu+1}} + \nu \geq 0$. Hence $E(y) \geq E(0) = 1$. Now take $\nu = a_2 + a_3 + a_4 + a_5$ and $y = 3r^{\frac{1}{2}} \lfloor \frac{t}{d'} \rfloor (R_0 - 1)$, and we get

$$\frac{1}{\left(1 + 3r^{\frac{1}{2}} \lfloor \frac{t}{d'} \rfloor (R_0 - 1)\right)^{a_2+a_3+a_4+a_5}} + 3(a_2 + a_3 + a_4 + a_5) r^{\frac{1}{2}} \lfloor \frac{t}{d'} \rfloor (R_0 - 1) \geq 1$$

The claim now follows immediately, as does the first statement of the theorem. If we are no longer restricted to the event $\{J_0 \leq \eta\}$, then (recall that T is the first time t such that $J_t \leq \eta$) by the remark (4.16)

$$\begin{aligned}
& \mathbb{P} [u^{t+3} \neq v^{t+3} | \mathcal{F}_0] \\
& \leq \mathbb{P} \left[u^{t+3} \neq v^{t+3} \mid J_0 > \eta, T \leq \left\lfloor \frac{t}{2} \right\rfloor + 3 \right] + \mathbb{P} \left[T > \left\lfloor \frac{t}{2} \right\rfloor + 3 \mid J_0 > \eta \right] \\
& \leq r^{\frac{1}{2} \lfloor \frac{t}{2d'} \rfloor} (1 + 3(a_2 + a_3 + a_4 + a_5)(R_0 - 1)) + \frac{\max\{J_0, \eta\} \beta^{\lfloor \frac{t}{2} \rfloor + 3}}{\eta} \quad (4.36)
\end{aligned}$$

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

4.8 Sampling from equilibrium

One application of the previously derived results lies in sampling from the equilibrium distribution π defined by (2.3). We will start by taking $\mathcal{U}^0 = (1, 1, 1, 1)$ and $\mathcal{V}^0 \sim \pi$, and define u_0, v_0 according to (4.7). Then from Theorem 4.2, using the one-shot coupling at time $t + 3$, it follows that for $t \geq d'$

$$\begin{aligned}
\mathbb{P} [u^{t+3} \neq v^{t+3} | \mathcal{F}_0] & \leq r^{\frac{1}{2} \lfloor \frac{t}{2d'} \rfloor} (1 + 3(a_2 + a_3 + a_4 + a_5)(R_0 - 1)) \\
& \quad + \frac{\max\{J_0, \eta\} \beta^{\lfloor \frac{t}{2} \rfloor + 3}}{\eta}
\end{aligned}$$

Proof of Corollary 4.3. This follows immediately from Theorem 4.2. \square

Now let $C_g := \int \left(\prod_{i=1}^4 z_i^{a_i+a_{i+1}-1} \right) \exp \left(\sum_{i=1}^5 -z_i z_{i-1} \right) dz$. Then we can bound the terms $\mathbb{E}_\pi [R_0]$ and $\mathbb{E}_\pi [J_0]$ in Corollary 4.3 in the following way:

$$\begin{aligned}
& d_{TV} (\mathcal{U}^{t+3}, \pi) \\
& \leq \mathbb{P} [u^{t+3} \neq v^{t+3}] \\
& \leq r^{\frac{1}{2} \lfloor \frac{t}{2d'} \rfloor} \frac{1}{C_g} (1 + 3a) \int \left(\frac{\max_i \{1, v_i\}}{\min_i \{1, v_i\}} \right) \left(\prod_{i=1}^4 v_i^{a_i+a_{i+1}-1} \right) \exp \left(\sum_{i=1}^5 -v_i v_{i-1} \right) dv \\
& + \frac{1}{\eta} \beta^{\lfloor \frac{t}{2} \rfloor + 3} \left(\eta + \frac{1}{C_g} \int J_0 \left(\prod_{i=1}^4 v_i^{a_i+a_{i+1}-1} \right) \exp \left(\sum_{i=1}^5 -v_i v_{i-1} \right) dv \right) \\
& \leq \tilde{C}_\pi r^{\frac{1}{2} \lfloor \frac{t}{2d'} \rfloor} (1 + 3a) + \left(\frac{\tilde{C}_J}{\eta} + 1 \right) \beta^{\lfloor \frac{t}{2} \rfloor + 3}
\end{aligned}$$

where $a = a_2 + a_3 + a_4 + a_5$,

$\tilde{C}_\pi := \int \left(\frac{\max_i \{1, v_i\}}{\min_i \{1, v_i\}} \right) \left(\prod_{i=1}^4 v_i^{a_i+a_{i+1}-1} \right) \exp \left(\sum_{i=1}^5 -v_i v_{i-1} \right) dv / C_g$, and

$\tilde{C}_J := \frac{1}{C_g} \int J_0 \left(\prod_{i=1}^4 v_i^{a_i+a_{i+1}-1} \right) \exp \left(\sum_{i=1}^5 -v_i v_{i-1} \right) dv$, and we derive a bound for these terms in Appendix B.

For the purpose of illustrating this result in a concrete example, let us set $x = 2$, $b = 3$ and $a_i = i$. Then by (26) $\tilde{C}_\pi \leq 60,300$ and by (27) $\tilde{C}_J \leq 59$, $\beta \leq 7/9$, $r \leq 1 - \frac{1}{2178}$, $20 \leq \eta \leq 21$ and $216 \leq d' \leq 217$, and hence

$$d_{TV} (\mathcal{U}^{t+3}, \pi) \leq 60300 * 43 \left(1 - \frac{1}{2178} \right)^{\frac{1}{2} \lfloor \frac{t}{434} \rfloor} + \left(1 + \frac{59}{20} \right) \left(\frac{7}{9} \right)^{\lfloor \frac{t}{2} \rfloor + 3}$$

which implies that $d_{TV} (\mathcal{U}^{t+3}, \pi) \leq 10^{-5}$ for $t \geq 50,000,000$.

4.9 A brief look at the case $n = 3$

The case $n = 3$ can be treated in a very similar manner as was done for $n = 4$. It follows immediately from (4.4) that this problem would reduce to dealing with a Markov chain of a single variable, given by

$$u^{t+1} = \frac{\gamma_2^{t+1}}{\frac{\gamma_1^{t+1}}{u^t+x} + \frac{\gamma_3^{t+1}}{u^t+b}} \quad (4.37)$$

Similarly, coupling two copies (u^t, v^t) uniformly with the property $u^0 \leq v^0$ implies that $u^t \leq v^t$. It is also an immediate observation that the ratio $R_t = \frac{v^t}{u^t}$ is strictly decreasing, hence we no longer need to define a process like (4.10) and can simply work with this ratio directly. It follows that $R_{t+1} = R_t Q_t$ where

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

where $\dot{r}_t := 1 - \min\{x, b\} / \left((u^t + \max\{x, b\}) \left(1 + \frac{\max\{x, b\}}{u^t}\right) \right)$. Note that if we define

$K_{1,t+1} := \frac{\gamma_2^{t+1}}{\gamma_1^{t+1} + \gamma_3^{t+1}} (u^t + x + b)$ and $K_{2,t+1} := \left(\frac{\gamma_1^{t+1}}{\gamma_2^{t+1}} \frac{1}{x} + \frac{\gamma_3^{t+1}}{\gamma_2^{t+1}} \frac{1}{b} \right)$ then $K_{1,t+1} \geq u^{t+1}$ and $K_{2,t+1} \geq \frac{1}{u^{t+1}}$, and hence we do not need a process analogous to D_t from the

previous section, since

$$\dot{r}_{t+1} \leq 1 - \min \{x, b\} / \left((K_{1,t+1} + \max \{x, b\}) (1 + \max \{x, b\} K_{2,t+1}) \right)$$

As before, we will require that $a_1 + a_4 > 1$ in order that $\mathbb{E} [\gamma_2 / (\gamma_1 + \gamma_3)] < 1$. If the process J_t and the stopping time S , as well as the constant η are also defined in an analogous manner, then we can repeat the steps of (4.19)

$$\begin{aligned} \mathbb{E} [R_{S+1}] &= \mathbb{E} [Q_S R_S] \\ &\leq \mathbb{E} [\dot{r}_S (R_S - 1)] + 1 \\ &= \mathbb{E} \left[\left(1 - \frac{\min \{x, b\}}{(u^S + \max \{x, b\}) \left(1 + \frac{\max \{x, b\}}{u^S} \right)} \right) (R_S - 1) \right] + 1 \\ &\leq \mathbb{E} \left[\left(1 - \frac{\min \{x, b\}}{(\eta + \max \{x, b\}) (1 + \eta \max \{x, b\})} \right) (R_S - 1) \right] + 1 \\ &= r \mathbb{E} [R_S - 1] + 1 \end{aligned} \tag{4.38}$$

where $r = 1 - \min \{x, b\} / \left((\eta + \max \{x, b\}) (1 + \eta \max \{x, b\}) \right)$. Note that we no longer need to look at time $S + 2$ in the left-hand side of (4.38) in order to obtain this inequality. This means that from the proof of Lemma 4.13 and Corollary 4.14 we get

$$\mathbb{E} [R_{t+1} | J_0 \leq \eta] \leq 1 + 3r \lfloor \frac{t}{\bar{x}} \rfloor (R_0 - 1)$$

From the proof of Theorem 4.2 we conclude

Theorem 4.15. $[n = 3]$ Suppose that $a_1 + a_4 > 1$. If u^t and v^t are two instances of the Markov chain (4.37)

$$d_{TV}(u^{t+2}, v^{t+2}) \leq r^{\lfloor \frac{t}{2d} \rfloor} (1 + 3(a_2 + a_3)(R_0 - 1)) + \frac{\max\{J_0, \eta\} \beta^{\lfloor \frac{t}{2} \rfloor + 3}}{\eta}$$

The requirement that $a_1 + a_4 > 1$ is mainly a necessity to make our proof work. However, it can be seen from Corollary 2.5 that under some conditions with $a_1 + a_4 \leq 1$, the Markov chain can lose some convergence properties.

We can make an analogous argument to obtain a result similar to Corollary 4.3. In particular if we let $\mathcal{U}^0 = (1, 1, 1)$, $\mathcal{V}^0 \sim \pi$ and $x = 1$, $b = 2$ and $a_i = i$, then by calculations similar to those done in Section 4.8 we get

$$d_{TV}(\mathcal{U}^{t+2}, \pi) \leq 600 \left(1 - \frac{1}{65}\right)^{\lfloor \frac{t}{100} \rfloor} + 6 \left(\frac{7}{9}\right)^{\lfloor \frac{t}{2} \rfloor + 3}$$

which in particular implies that $d_{TV}(\mathcal{U}^{t+2}, \pi) \leq 10^{-5}$ for $t \geq 125,000$. This is fewer iterations than what was established for the Wasserstein bound in Section 2.1, and with some effort it may be possible to translate the above d_{TV} bound into a better d_W bound than what is given in that section. Nevertheless, the analysis in Section 2.1 gives a very different methodology that may be of interest to many readers.

4.10 $n > 4$

It is not difficult to show that the method outlined in (4.10) and leading to inequality (4.11) can also be extended to the case where $n = 5$: given two starting points $u^0 \preceq v^0 \in \mathbb{R}_+^5$, it amounts to setting $u^{t+1} := f^{t+1}(u^t)$ and $v^{t+1} := \lambda_{t+1} u^{t+1}$, where $\lambda_{t+1} := \max \left\{ \frac{f_2^{t+1}(v^t)}{u_2^{t+1}}, \frac{f_4^{t+1}(v^t)}{u_4^{t+1}} \right\}$. A calculation similar to (4.11) shows that $R_{t+1} = \lambda_{t+1} < R_t$ whenever $u^t \neq v^t$. If we attempt to replicate this method for $n \geq 6$ however, it becomes apparent that $\lambda_t = R_0$, so that $R_t = R_0$ is fixed for all times t . It is nonetheless possible to extend this method for $n \geq 6$ if we consider a multi-step version of the Markov chain in (4.4). More precisely, let $f_m : \mathbb{R}_+^m \rightarrow \mathbb{R}_+^m$ be defined by

$$\begin{aligned} f_m(u_2, u_4, \dots, u_{2m}) &:= \left(\frac{\gamma_2}{\frac{\gamma_1}{x+u_2} + \frac{\gamma_3}{u_2+u_4}}, \frac{\gamma_4}{\frac{\gamma_3}{u_2+u_4} + \frac{\gamma_5}{u_4+u_6}}, \dots, \frac{\gamma_{2m}}{\frac{\gamma_{2m-1}}{u_{2(m-1)}+u_{2m}} + \frac{\gamma_{2m+1}}{u_{2m}+b}} \right) \\ &= (f_{(m,2)}(u), f_{(m,4)}(u), \dots, f_{(m,2m)}(u)) \end{aligned}$$

and let

$$F_m^k := f_m^k \circ f_m^{k-1} \circ \dots \circ f_m^1 \quad (4.39)$$

Lemma 4.16. *Suppose $u \preceq v$ and $\frac{v_2}{u_2} = \frac{v_4}{u_4} = \dots = \frac{v_{2m}}{u_{2m}}$. Then for $m \geq 3$ and*

$$j \in \left\{ 1, \dots, \left\lceil \frac{m}{2} \right\rceil - 1, \left\lceil \frac{m+1}{2} \right\rceil + 1, \dots, m \right\}$$

$$\frac{F_{(m,2j)}^{\left\lceil \frac{m}{2} \right\rceil - 1}(v)}{F_{(m,2j)}^{\left\lceil \frac{m}{2} \right\rceil - 1}(u)} < \frac{v_2}{u_2} \quad (4.40)$$

and

$$\frac{F_{(m,2\lceil \frac{m}{2} \rceil)}^{\lceil \frac{m}{2} \rceil - 1}(v)}{F_{(m,2\lceil \frac{m}{2} \rceil)}^{\lceil \frac{m}{2} \rceil - 1}(u)} = \frac{F_{(m,2\lceil \frac{m+1}{2} \rceil)}^{\lceil \frac{m}{2} \rceil - 1}(v)}{F_{(m,2\lceil \frac{m+1}{2} \rceil)}^{\lceil \frac{m}{2} \rceil - 1}(u)} = \frac{v_2}{u_2} \quad (4.41)$$

Furthermore, for $j \in \{1, \dots, m\}$ we get

$$\lambda := \frac{F_{(m,2j)}^{\lceil \frac{m}{2} \rceil}(v)}{F_{(m,2j)}^{\lceil \frac{m}{2} \rceil}(u)} < \frac{v_2}{u_2} \quad (4.42)$$

We could hence consider the Markov chain defined by the i.i.d. random functions $\{h_t\}$, where $h_t \sim F_m^{\lceil \frac{m}{2} \rceil}$, and by (4.42) it would be guaranteed that $\lambda_t = R_{t+1} < R_t$. A proof of this lemma is given in Appendix A. It is also implied here that $n = 2m+1$ is odd, which bears no impact on the argument in the proof, as it would be identical for n even.

Despite this contraction of R_t implied by Lemma 4.16, it is not a straight forward matter to show (and derive an upper bound for) the geometric convergence of $R_t \rightarrow 1$, primarily due to the difficulty of handling the 'continued fraction' form of the functions h_t . We will therefore consider an alternative Markov chain. For odd $n = 2m + 1$ we will define this by

$$\begin{aligned} (u_2^{t+1}, u_4^{t+1}, \dots, u_{2m}^{t+1}) &= g(u_2^t, u_4^t, \dots, u_{2m}^t) \\ &:= \left(\frac{\gamma_2}{\frac{\tilde{\gamma}_1}{x+u_2^t} + \frac{\gamma_3}{u_2^t+u_4^t}}, \frac{\gamma_4}{\frac{\tilde{\gamma}_3}{u_2^{t+1}+u_4^t} + \frac{\gamma_5}{u_4^t+u_6^t}}, \dots, \frac{\gamma_{2m}}{\frac{\tilde{\gamma}_{2m-1}}{u_{2m-2}^{t+1}+u_{2m}^t} + \frac{\gamma_{2m+1}}{u_{2m}^t+b}} \right) \end{aligned} \quad (4.43)$$

where $\{\gamma_2, \gamma_3, \dots, \gamma_{2n+1}\}$ are same as before and $\{\tilde{\gamma}_1, \tilde{\gamma}_3, \dots, \tilde{\gamma}_{2n-1}\}$ is an i.i.d. copy of $\{\gamma_1, \gamma_3, \dots, \gamma_{2n-1}\}$. The definition of g for n even is as one might expect. Furthermore it will be apparent that all arguments for n even would be non-distinct from ones about to be made for odd n , which is why we shall forgo the separate treatment of this case.

Note also that the random function in (4.43) can be identified with the transition kernel

$$\bar{O} := (\bar{P}_1 \bar{P}_3 \bar{P}_5 \dots \bar{P}_{2m+1}) (\bar{P}_2 \bar{P}_3 \bar{P}_4 \dots \bar{P}_{2m}) \quad (4.44)$$

The probability kernel \bar{O} appears similar to \bar{P} defined in (4.1). There is however no obvious way to state a clear relation between the two. We observe that the kernel $(\bar{P}_1 \bar{P}_3 \bar{P}_5 \dots \bar{P}_{2m+1})$ is responsible for generating the variables $\{\tilde{\gamma}_1, \gamma_3, \gamma_5, \dots, \gamma_{2m+1}\}$ while $(\bar{P}_2 \bar{P}_3 \bar{P}_4 \dots \bar{P}_{2m})$ generates $\{\gamma_2, \tilde{\gamma}_3, \gamma_4, \tilde{\gamma}_5, \dots, \tilde{\gamma}_{2m-1}, \gamma_{2m}\}$. Recalling that $\pi \bar{P}_i = \pi$, it is evident that π is also invariant with respect to \bar{O} , which shows that this Markov chain will also converge to π in distribution (it is not difficult to ascertain that this is indeed a Harris chain).

We can now extend the method we used for the case $n = 4$ to general n , with the Markov chain defined by the random functions $\{g^t\}$: starting with $u^0 \preceq v^0$

such that $\frac{v_2^0}{u_2^0} = \dots = \frac{v_{2m}^0}{u_{2m}^0} > 1$, set $u^{t+1} := g^{t+1}(u^t)$ and $v^{t+1} := R_{t+1}u^{t+1}$ where

$$R_{t+1} := \max_j \left\{ \frac{g_{2j}^{t+1}(v^t)}{g_{2j}^{t+1}(u^t)} \right\} \quad (4.45)$$

$$= \frac{v_2^t}{u_2^t} \max \left\{ \frac{\frac{\tilde{\gamma}_1^{t+1}}{1+\frac{x}{u_2^t}} + \frac{\gamma_3^{t+1}}{1+\frac{u_4^t}{u_2^t}}}{\frac{\tilde{\gamma}_1^{t+1}}{1+\frac{x}{v_2^t}} + \frac{\gamma_3^{t+1}}{1+\frac{v_4^t}{v_2^t}}}, \frac{\frac{\tilde{\gamma}_3^{t+1}}{1+\frac{g_{2j}^{t+1}(u^t)}{u_4^t}} + \frac{\gamma_5^{t+1}}{1+\frac{u_6^t}{u_4^t}}}{\frac{\tilde{\gamma}_3^{t+1}}{1+\frac{g_{2j}^{t+1}(v^t)}{v_4^t}} + \frac{\gamma_5^{t+1}}{1+\frac{v_6^t}{v_4^t}}}, \dots, \frac{\frac{\tilde{\gamma}_{2m-1}^{t+1}}{1+\frac{g_{2m-2}^{t+1}(u^t)}{u_{2m}^t}} + \frac{\gamma_{2m+1}^{t+1}}{1+\frac{b}{u_{2m}^t}}}{\frac{\tilde{\gamma}_{2m-1}^{t+1}}{1+\frac{g_{2m-2}^{t+1}(v^t)}{v_{2m}^t}} + \frac{\gamma_{2m+1}^{t+1}}{1+\frac{b}{v_{2m}^t}}} \right\}$$

Here we have used the notation $g = (g_2, g_4, \dots, g_{2m})$ to represent the components of the function g . In (4.45) we have extracted the factor $\frac{v_2^t}{u_2^t}$ by making the implicit assumption that $\frac{v_2^t}{u_2^t} = \frac{v_4^t}{u_4^t} = \dots = \frac{v_{2m}^t}{u_{2m}^t}$. The validity of this is evident from the definition of the process v^t , and inductively (in t) from (4.45). Let $R_0 = \frac{v_0^0}{u_0^0}$. We

can then confirm by a simple inductive argument that $\frac{g_{2j}^{t+1}(v^t)}{g_{2j}^{t+1}(u^t)} < R_t$ for all j and $t \geq 0$ as follows: it is immediate that $\frac{g_2^{t+1}(v^t)}{g_2^{t+1}(u^t)} = R_t \left(\frac{\frac{\tilde{\gamma}_1^{t+1}}{1+\frac{x}{u_2^t}} + \frac{\gamma_3^{t+1}}{1+\frac{u_4^t}{u_2^t}}}{\frac{\tilde{\gamma}_1^{t+1}}{1+\frac{x}{v_2^t}} + \frac{\gamma_3^{t+1}}{1+\frac{v_4^t}{v_2^t}}} \right) < R_t$, since $\frac{u_4^t}{u_2^t} = \frac{v_4^t}{v_2^t}$ while $\frac{x}{u_2^t} > \frac{x}{v_2^t}$. Now assuming that $\frac{g_{2j}^{t+1}(v^t)}{g_{2j}^{t+1}(u^t)} < R_t$, we get

$$\frac{g_{2j}^{t+1}(u^t)}{u_{2j+2}^t} > \frac{g_{2j}^{t+1}(v^t)}{v_{2j+2}^t} \quad (4.46)$$

(since $\frac{v_{2j+2}^t}{u_{2j+2}^t} = R_t$), hence $\frac{g_{2j+2}^{t+1}(v^t)}{g_{2j+2}^{t+1}(u^t)} = R_t \left(\frac{\frac{\tilde{\gamma}_{2j+1}^{t+1}}{1+\frac{g_{2j}^{t+1}(u^t)}{u_{2j+2}^t}} + \frac{\gamma_{2j+3}^{t+1}}{1+\frac{u_{2j+4}^t}{u_{2j+2}^t}}}{\frac{\tilde{\gamma}_{2j+1}^{t+1}}{1+\frac{g_{2j}^{t+1}(v^t)}{v_{2j+2}^t}} + \frac{\gamma_{2j+3}^{t+1}}{1+\frac{v_{2j+4}^t}{v_{2j+2}^t}}} \right) < R_t$ (again since $\frac{u_{2j+4}^t}{u_{2j+2}^t} \geq \frac{v_{2j+4}^t}{v_{2j+2}^t}$, where by convention we take $u_{2m+2}^t = v_{2m+2}^t = b$), which completes

this inductive argument.

Let us now consider the i^{th} term in the right-hand side of (4.45). If we replace

both

$\gamma_{2i+1}^{t+1}/(1+u_{2i+2}^t/u_{2i}^t)$ in the numerator and $\gamma_{2i+1}^{t+1}/(1+v_{2i+2}^t/v_{2i}^t)$ in the denominator by γ_{2i+1}^{t+1} , then by Lemma 4.8 the right-hand side of (4.45) would not decrease.

Hence we can say that

$$\begin{aligned}
& R_{t+1} \\
& \leq \frac{v_2^t}{u_2^t} \max \left\{ \frac{\frac{\tilde{\gamma}_1^{t+1}}{1+\frac{x}{u_2^t}} + \gamma_3^{t+1}}{\frac{\tilde{\gamma}_1^{t+1}}{1+\frac{x}{v_2^t}} + \gamma_3^{t+1}}, \frac{\frac{\tilde{\gamma}_3^{t+1}}{1+\frac{g_2^{t+1}(u^t)}{u_4^t}} + \gamma_5^{t+1}}{\frac{\tilde{\gamma}_3^{t+1}}{1+\frac{g_2^{t+1}(v^t)}{v_4^t}} + \gamma_5^{t+1}}, \dots, \frac{\frac{\tilde{\gamma}_{2m-1}^{t+1}}{1+\frac{g_{2m-2}^{t+1}(u^t)}{u_{2m}^t}} + \gamma_{2m+1}^{t+1}}{\frac{\tilde{\gamma}_{2m-1}^{t+1}}{1+\frac{g_{2m-2}^{t+1}(v^t)}{v_{2m}^t}} + \gamma_{2m+1}^{t+1}} \right\} \quad (4.47) \\
& =: \frac{v_2^t}{u_2^t} Q_{t+1}
\end{aligned}$$

We can proceed in a manner similar to what we did in Lemma 4.9. Let $\mu_1 := a_3 + a_4 + \dots + a_{2m+2}$ and $\mu_2 := \frac{1}{a_1+a_2-1} + \dots + \frac{1}{a_{2m-1}+a_{2m}-1}$. Then by [9] and Markov's inequality

$$\begin{aligned}
& \mathbb{P} \left[\bigcap_{i \text{ odd}, i=1}^{2m-1} \left\{ \tilde{\gamma}_i \geq \frac{1}{2\mu_2} \right\}, \bigcap_{i \text{ odd}, i=3}^{2m+1} \left\{ \gamma_i \leq \mu_1 \right\} \right] \\
& = \mathbb{P} \left[\bigcap_{i \text{ odd}, i=1}^{2m-1} \left\{ \frac{1}{\tilde{\gamma}_i} < 2\mu_2 \right\} \right] \mathbb{P} \left[\bigcap_{i \text{ odd}, i=3}^{2m+1} \left\{ \gamma_i \leq \mu_1 \right\} \right] \\
& \geq \mathbb{P} \left[\sum_{i \text{ odd}, i=1}^{2m-1} \frac{1}{\tilde{\gamma}_i} < 2\mu_2 \right] \mathbb{P} \left[\sum_{i \text{ odd}, i=3}^{2m+1} \gamma_i \leq \mu_1 \right] \\
& \geq \frac{1}{4} \quad (4.48)
\end{aligned}$$

Lemma 4.17. Let $M_{t+1} := \max_{2 \leq i \leq m} \left\{ 2 + \frac{g_{2i-2}^{t+1}(v^t)}{v_{2i}^t} + \frac{u_{2i}^t}{u_{2i-2}^{t+1}} \right\}$ and

$$\dot{r}_{t+1} := 1 - \frac{1}{(1+2\mu_1\mu_2)^m (M_{t+1})^{m-1} \left(1 + \frac{x}{v_2^t}\right) \left(1 + \frac{u_2^t}{x}\right)}. \text{ Then}$$

$$\mathbb{E}[R_{t+1}] \leq \mathbb{E} \left[\left(\frac{1}{4} \dot{r}_{t+1} + \frac{3}{4} \right) (R_t - 1) \right] + 1$$

Proof. Let $\dot{r}_{t+1,2} := 1 - \frac{1}{(1+2\mu_1\mu_2) \left(1 + \frac{x}{v_2^t}\right) \left(1 + \frac{u_2^t}{x}\right)}$, and define recursively $\dot{r}_{t+1,2i} := 1 - \frac{(1-\dot{r}_{t+1,2i-2})}{(1+2\mu_1\mu_2)M_{t+1}}$ for $2 \leq i \leq m$. Let $Q_{t+1,i}$ be the i^{th} term inside the max in (4.47).

We claim that in the event

$\left\{ \bigcap_{i \text{ odd}, i=1}^{2m-1} \left\{ \tilde{\gamma}_i^{t+1} \geq \frac{1}{2\mu_2} \right\}, \bigcap_{i \text{ odd}, i=3}^{2m+1} \left\{ \gamma_i^{t+1} \leq \mu_1 \right\} \right\}$, the term $Q_{t+1,i}$ is bounded from above by $\dot{r}_{t+1,2i} + \frac{1-\dot{r}_{t+1,2i}}{R_t}$ for each i . Assume that this statement is true for $i-1$. Note that since the $i-1^{\text{st}}$ term inside the *max* in (4.45) is less than or equal to the $i-1^{\text{st}}$ term inside the *max* in (4.47), this implies that $\frac{g_{2i-2}^{t+1}(v^t)}{g_{2i-2}^{t+1}(u^t)} \leq \frac{v_2^t}{u_2^t} Q_{t+1,i-1} \leq R_t \left(\dot{r}_{t+1,2i-2} + \frac{1-\dot{r}_{t+1,2i-2}}{R_t} \right)$. Then by Lemma 4.8 and the fact that

$\frac{g_{2i-2}^{t+1}(u^t)}{u_{2i}^t} > \frac{g_{2i-2}^{t+1}(v^t)}{v_{2i}^t}$ (which follows from (4.46)), we get

$$\begin{aligned}
& \frac{\frac{\tilde{\gamma}_{2i-1}^{t+1}}{1 + \frac{g_{2i-2}^{t+1}(u^t)}{u_{2i}^t}} + \gamma_{2i+1}^{t+1}}{\frac{\tilde{\gamma}_{2i-1}^{t+1}}{1 + \frac{g_{2i-2}^{t+1}(v^t)}{v_{2i}^t}} + \gamma_{2i+1}^{t+1}}}}{\frac{\frac{1}{1 + \frac{g_{2i-2}^{t+1}(u^t)}{u_{2i}^t}} + \mu_1}{\frac{1}{1 + \frac{g_{2i-2}^{t+1}(v^t)}{v_{2i}^t}} + \mu_1}}} = 1 - \frac{\frac{1}{1 + \frac{g_{2i-2}^{t+1}(v^t)}{v_{2i}^t}} - \frac{1}{1 + \frac{g_{2i-2}^{t+1}(u^t)}{u_{2i}^t}}}{\frac{1}{1 + \frac{g_{2i-2}^{t+1}(v^t)}{v_{2i}^t}} + 2\mu_1\mu_2}}{1 - \frac{g_{2i-2}^{t+1}(v^t)}{v_{2i}^t} \frac{u_{2i}^t}{g_{2i-2}^{t+1}(u^t)}}} \\
& = 1 - \frac{\left(\frac{1}{1 + \frac{g_{2i-2}^{t+1}(v^t)}{v_{2i}^t}} + 2\mu_1\mu_2 \right) \left(1 + \frac{g_{2i-2}^{t+1}(v^t)}{v_{2i}^t} \right) \left(1 + \frac{g_{2i-2}^{t+1}(u^t)}{u_{2i}^t} \right) \frac{u_{2i}^t}{g_{2i-2}^{t+1}(u^t)}}{\left(1 - \frac{1}{R_t} \left(R_t \left(\dot{r}_{t+1,2i-2} + \frac{1-\dot{r}_{t+1,2i-2}}{R_t} \right) \right) \right)} \\
& \leq 1 - \frac{\left(1 - \frac{1}{R_t} \left(R_t \left(\dot{r}_{t+1,2i-2} + \frac{1-\dot{r}_{t+1,2i-2}}{R_t} \right) \right) \right)}{\left(1 + 2\mu_1\mu_2 \right) \left(1 + \frac{g_{2i-2}^{t+1}(v^t)}{v_{2i}^t} \right) \left(1 + \frac{u_{2i}^t}{g_{2i-2}^{t+1}(u^t)} \right)} \\
& \leq 1 - \frac{\left(1 - \dot{r}_{t+1,2i-2} \right) \left(1 - \frac{1}{R_t} \right)}{\left(1 + 2\mu_1\mu_2 \right) \left(2 + \frac{g_{2i-2}^{t+1}(v^t)}{v_{2i}^t} + \frac{u_{2i}^t}{g_{2i-2}^{t+1}(u^t)} \right)} \\
& \leq \dot{r}_{t+1,2i} + \frac{1 - \dot{r}_{t+1,2i}}{R_t}
\end{aligned}$$

This proves the inductive step. The computation showing $Q_{t+1,1} \leq \dot{r}_{t+1,2} + \frac{1-\dot{r}_{t+1,2}}{R_t}$

is identical to the one given above. Observe that $\dot{r}_{t+1,2} \leq \dot{r}_{t+1,4} \leq \dots \leq \dot{r}_{t+1,2m}$, or

more precisely

$$\begin{aligned}
1 - \dot{r}_{t+1,2i} &= \frac{(1 - \dot{r}_{t+1,2i-2})}{(1 + 2\mu_1\mu_2) M_{t+1}} \\
&= \left(\frac{1}{(1 + 2\mu_1\mu_2) M_{t+1}} \right)^{i-1} (1 - \dot{r}_{t+1,2}) \\
&= \frac{1}{(1 + 2\mu_1\mu_2)^i (M_{t+1})^{i-1} \left(1 + \frac{x}{v_2^t}\right) \left(1 + \frac{u_2^t}{x}\right)}
\end{aligned} \tag{4.49}$$

We conclude that $Q_{t+1} \leq \dot{r}_{t+1,2m} + \frac{1 - \dot{r}_{t+1,2m}}{R_t}$, and hence by (4.48)

$$\mathbb{E}[R_{t+1}] \leq \mathbb{E} \left[\left(\frac{1}{4} \dot{r}_{t+1} + \frac{3}{4} \right) (R_t - 1) \right] + 1$$

□

An apparent weakness of (4.49) is that the bound on the rate of convergence is exponentially bad in n . Unfortunately, this is an inherent property of this method: observe that even under the most favourable scenario in (4.45) (whereby we take $\gamma_3 = \gamma_5 = \dots = \gamma_{2n+1} = 0$), one still arrives at

$$\begin{aligned}
\dot{r}_{t+1,2} &= 1 - \frac{1}{1 + \frac{u_2^t}{x}} \\
\dot{r}_{t+1,2i} &= 1 - \frac{(1 - \dot{r}_{t+1,2i-2})}{1 + \frac{u_{2i}^t}{u_{2i-2}^{t+1}}} \\
&\geq 1 - \left(\frac{1}{m_{t+1}} \right)^i
\end{aligned}$$

where $m_{t+1} = \min_i \left\{ 1 + \frac{u_{2i}^t}{u_{2i-2}^{t+1}} \right\}$.

Note that the results derived in section 2.1.1 are in fact independent of many aspects of the model, including the dimension n as well as the Markov chain in

question. They relied only on the existence of auxiliary processes $(K_{i,t}, J_t, D_t)$, stochastic monotonicity of the two paths, and the existence of a non-increasing process R_t as described in the prelude to Lemma 4.7. At this point we only require to ascertain the existence of an auxiliary processes $\{K_t\}$ that satisfies conditions similar to the ones we had for the case $n = 4$, and which assists in bounding M_t from above. We will prove such existence under certain restrictions.

4.11 Construction of $\{K_{i,t}\}$ for $n \geq 5$

For $i = 1, \dots, m$ let $w_{2i}^t := \frac{1}{w_{2i}^t}$. Then

$$w_{2i}^{t+1} = \frac{1}{\gamma_{2i}^{t+1}} \left(\frac{\tilde{\gamma}_{2i-1}^{t+1}}{\frac{1}{w_{2i-2}^{t+1}} + \frac{1}{w_{2i}^t}} + \frac{\gamma_{2i+1}^{t+1}}{\frac{1}{w_{2i}^t} + \frac{1}{w_{2i+2}^t}} \right)$$

where for convenience we have taken $w_0^t = \frac{1}{x}$ and $w_{2m+2}^t = \frac{1}{b}$ for all t . We will make use of the following inequality: for $a, b, \rho_1, \rho_2 \in \mathbb{R}^+$,

$$\left(\frac{a}{\rho_1} + \frac{b}{\rho_2} \right) \geq \frac{(a+b)^2}{(a\rho_1 + b\rho_2)} \quad (4.50)$$

Hence, for $1 \leq i \leq m$

$$\begin{aligned} w_{2i}^{t+1} &= \frac{1}{\gamma_{2i}^{t+1}} \left(\frac{\tilde{\gamma}_{2i-1}^{t+1}}{\frac{1}{w_{2i-2}^{t+1}} + \frac{1}{w_{2i}^t}} + \frac{\gamma_{2i+1}^{t+1}}{\frac{1}{w_{2i}^t} + \frac{1}{w_{2i+2}^t}} \right) \\ &\leq \frac{1}{\gamma_{2i}^{t+1}} \left(\frac{\tilde{\gamma}_{2i-1}^{t+1}}{4} (w_{2i-2}^{t+1} + w_{2i}^t) + \frac{\gamma_{2i+1}^{t+1}}{4} (w_{2i}^t + w_{2i+2}^t) \right). \end{aligned} \quad (4.51)$$

We can now exploit the linearity in (4.51) to get an upper bound on $\mathbb{E}[K_{1,t+1} | \mathcal{F}_t]$ where $K_{1,t} := \sum_{i=1}^m w_{2i}^t$. Observe that

$$\begin{aligned} & \mathbb{E}[K_{1,t+1} | \mathcal{F}_t] \\ &= \sum_{i=1}^m \mathbb{E}[w_{2i}^{t+1} | \mathcal{F}_t] \\ &\leq \sum_{i=1}^m \frac{1}{\alpha_{2i} - 1} \left(\frac{\alpha_{2i-1}}{4} \left(\mathbb{E}[w_{2i-2}^{t+1} | \mathcal{F}_t] + w_{2i}^t \right) + \frac{\alpha_{2i+1}}{4} (w_{2i}^t + w_{2i+2}^t) \right) \end{aligned} \quad (4.52)$$

We will re-write the right-hand side of (4.52) in a form that will reduce it to a super-martingale type of inequality, analogous to (4.32) for the $n = 4$ case. Let $A_i = \mathbb{E}[w_{2i}^{t+1} | \mathcal{F}_t]$, $B_i = w_{2i}^t$ for $1 \leq i \leq m$, and $A_0 = B_0 = \frac{1}{x}$ and $A_{m+1} = B_{m+1} = \frac{1}{b}$. Let $C_i^+ = \frac{\alpha_{2i+1}}{4(\alpha_{2i}-1)}$, $C_i^- = \frac{\alpha_{2i-1}}{4(\alpha_{2i}-1)}$ and $D_i = C_i^+ + C_i^-$ for $1 \leq i \leq m$, and $C_0^+ = C_0^- = 0$. Then by (4.51), for $1 \leq i \leq m$

$$A_i \leq C_i^- A_{i-1} + D_i B_i + C_i^+ B_{i+1} \quad (4.53)$$

In particular, since $A_0 = \frac{1}{x} = B_0$

$$A_1 \leq C_1^- B_0 + D_1 B_1 + C_1^+ B_2 \quad (4.54)$$

Define $q_{i,j}$ as follows: $q_{1,0} = C_1^-$, $q_{1,1} = D_1$, $q_{1,2} = C_1^+$, and $q_{1,j} = 0$ for $j > 2$ (so

that $A_1 \leq \sum_{j=0}^{m+1} q_{1,j} B_j$ by (4.54)); and for $2 \leq i \leq m$,

$$\begin{aligned} q_{i,i+1} &= C_i^+ \\ q_{i,i} &= C_i^- q_{i-1,i} + D_i \\ q_{i,j} &= C_i^- q_{i-1,j} \quad \text{for } 0 \leq j < i \\ q_{i,j} &= 0 \quad \text{for } j > i + 1 \end{aligned}$$

Then for $2 \leq i \leq m$

$$A_i \leq \sum_{j=0}^{m+1} q_{i,j} B_j$$

which follows from (4.53) and (4.54) and by induction on i .

Then the formulas for $q_{i,j}$ ($1 \leq i \leq m$) are:

$$q_{i,0} = \prod_{k=1}^i C_k^-$$

and for $1 \leq j \leq m + 1$

$$\begin{aligned} q_{i,j} &= 0 \quad \text{for } i < j - 1 \\ q_{j-1,j} &= C_{j-1}^+ \\ q_{j,j} &= C_j^- q_{j-1,j} + D_j = C_j^- C_{j-1}^+ + D_j \\ q_{i,j} &= C_i^- q_{i-1,j} = \left(C_j^- C_{j-1}^+ + D_j \right) \prod_{k=j+1}^i C_k^- \quad \text{for } i > j \end{aligned}$$

Next, let

$$\zeta_{2j} = \sum_{i=1}^m q_{i,j} \tag{4.55}$$

Then

$$\mathbb{E} [K_{1,t+1} | \mathcal{F}_t] = \sum_{i=1}^m A_i \leq \sum_{j=0}^{m+1} \zeta_{2j} B_j = \sum_{j=0}^{m+1} \zeta_{2j} w_{2i}^t \quad (4.56)$$

We have $\zeta_{2m+2} = q_{m,m+1} = C_m^+$ and $\zeta_0 = \sum_{i=1}^m \prod_{k=1}^i C_k^-$. For $1 \leq j \leq m$, we have

$$\begin{aligned} \zeta_{2j} &= \sum_{i=j-1}^m q_{i,j} \\ &= C_{j-1}^+ + \left(C_j^- C_{j-1}^+ + D_j \right) \left(\sum_{i=j}^m \prod_{k=j+1}^i C_k^- \right) \end{aligned} \quad (4.57)$$

We can obtain a similar result for $\mathbb{E} [K_{2,t} | \mathcal{F}_t]$ where $K_{2,t} := \sum_{i=1}^m u_{2i}^t$. Setting $u_0^t = x$ and $u_{n+1}^t = b$ for all t , it follows from (4.50) that for $1 \leq i \leq m$

$$\begin{aligned} u_{2i}^{t+1} &= \frac{\gamma_{2i}^{t+1}}{\frac{\tilde{\gamma}_{2i-1}^{t+1}}{u_{2i-2}^{t+1} + u_{2i}^t} + \frac{\gamma_{2i+1}^{t+1}}{u_{2i}^t + u_{2i+2}^t}} \\ &\leq \gamma_{2i}^{t+1} \left(\frac{\tilde{\gamma}_{2i-1}^{t+1}}{(\tilde{\gamma}_{2i-1}^{t+1} + \gamma_{2i+1}^{t+1})^2} (u_{2i-2}^{t+1} + u_{2i}^t) + \frac{\gamma_{2i+1}^{t+1}}{(\tilde{\gamma}_{2i-1}^{t+1} + \gamma_{2i+1}^{t+1})^2} (u_{2i}^t + u_{2i+2}^t) \right) \end{aligned}$$

Since $\frac{1}{(\tilde{\gamma}_{2i-1}^{t+1} + \gamma_{2i+1}^{t+1})^2}$ is a decreasing function of $\tilde{\gamma}_{2i-1}^{t+1}$, by Harris' inequality (page 136

of [2]) we get

$$\mathbb{E} \left[\frac{\tilde{\gamma}_{2i-1}^{t+1}}{(\tilde{\gamma}_{2i-1}^{t+1} + \gamma_{2i+1}^{t+1})^2} \right] \leq \frac{\alpha_{2i-1}}{(\alpha_{2i-1} + \alpha_{2i+1} - 1)(\alpha_{2i-1} + \alpha_{2i+1} - 2)}. \text{ Therefore, in an analogous man-}$$

ner to the previous derivations, we let $E_i = \mathbb{E} [u_{2i}^{t+1} | \mathcal{F}_t]$, $F_i = u_{2i}^t$ for $1 \leq i \leq m$,

and $E_0 = F_0 = x$ and $E_{m+1} = F_{m+1} = b$. Let $G_i^+ = \frac{\alpha_{2i} \alpha_{2i+1}}{(\alpha_{2i-1} + \alpha_{2i+1} - 1)(\alpha_{2i-1} + \alpha_{2i+1} - 2)}$,

$G_i^- = \frac{\alpha_{2i} \alpha_{2i-1}}{(\alpha_{2i-1} + \alpha_{2i+1} - 1)(\alpha_{2i-1} + \alpha_{2i+1} - 2)}$ and $H_i = G_i^+ + G_i^-$ for $1 \leq i \leq m$, and $G_0^+ =$

$G_0^- = 0$. We also define $p_{i,j}$ in an analogous manner to $q_{i,j}$ so that $E_i \leq \sum_{j=0}^{m+1} p_{ij} F_j$,

and let

$$\begin{aligned}\xi_{2j} &= \sum_{i=1}^m p_{i,j} \\ &= G_{j-1}^+ + \left(G_j^- G_{j-1}^+ + D_j\right) \left(\sum_{i=j}^m \prod_{k=j+1}^i H_k^-\right)\end{aligned}\tag{4.58}$$

Then

$$\mathbb{E} [K_{2,t+1} | \mathcal{F}_t] = \sum_{i=1}^m E_i \leq \sum_{j=0}^{m+1} \xi_{2j} F_j = \sum_{j=0}^{m+1} \xi_{2j} u_{2j}^t\tag{4.59}$$

It is now immediate that $\mathbb{E} [K_{2,t+1} | \mathcal{F}_t] \leq \max \{\xi_i\} K_{2,t} + C_2$ and $\mathbb{E} [K_{1,t+1} | \mathcal{F}_t] \leq \max \{\zeta_i\} K_{1,t} + C_1$ where $C_1 := \zeta_0 \frac{1}{x} + \zeta_{2m+2} \frac{1}{m}$ and $C_2 := \xi_0 x + \xi_{2m+2} m$, which was the goal of the last derivations.

We can now repeat the argument that led to (4.19), to obtain the following analogous inequality whenever $\max \{\zeta_i, \xi_i\} < 1$. Let the stopping time S be adapted to \mathcal{F}_t such that $J_S := K_{1,S} + K_{2,S} \leq \eta$, where $\eta := \frac{2(C_1+C_2)}{1-\max\{\zeta_2, \dots, \zeta_{2m}, \xi_2, \dots, \xi_{2m}\}}$. Furthermore, observe that for $1 \leq i \leq 2m$

$$\begin{aligned}K_{1,t} K_{2,t+1} + K_{1,t+1} K_{2,t} &\geq \frac{u_{2i-2}^{t+1}}{u_{2i}^t} + \frac{u_{2i}^t}{u_{2i-2}^{t+1}} + \frac{u_{2i}^{t+1}}{u_{2i-2}^t} + \frac{u_{2i-2}^t}{u_{2i}^{t+1}} \\ &\geq \frac{u_{2i-2}^{t+1}}{u_{2i}^t} + \frac{u_{2i}^t}{u_{2i-2}^{t+1}} + 2\end{aligned}\tag{4.60}$$

Therefore, setting $\mu := 1 + 2\mu_1\mu_2$, substituting (4.60) in the definition of M_{t+1} and

using the fact $\frac{g_{2i-2}^{t+1}(v^t)}{v_{2i}^t} \leq \frac{u_{2i-2}^{t+1}}{u_{2i}^t}$ we get

$$\begin{aligned}
\mathbb{E}[R_{S+1}] &\leq \mathbb{E} \left[\left(\frac{1}{4} \dot{r}_{S+1,2m} + \frac{3}{4} \right) (R_S - 1) \right] + 1 \\
&\leq \mathbb{E} \left[\left(1 - \frac{1}{4} \frac{\left((1 + 2\mu_1\mu_2) \left(1 + \frac{x}{v_2^S} \right) \left(1 + \frac{u_2^S}{x} \right) \right)^{-1}}{\left((1 + 2\mu_1\mu_2) M_{S+1} \right)^{m-1}} \right) (R_S - 1) \right] + 1 \\
&\leq \mathbb{E} \left[\left(1 - \frac{1}{4} \frac{\left((1 + 2\mu_1\mu_2) \left(1 + xK_{1,S} \right) \left(1 + \frac{K_{2,S}}{x} \right) \right)^{-1}}{\left((1 + 2\mu_1\mu_2) \left(K_{1,S+1}K_{2,S} + K_{2,S+1}K_{1,S} \right) \right)^{m-1}} \right) (R_S - 1) \right] + 1 \\
&\leq \mathbb{E} \left[\left(1 - \frac{1}{4} \frac{\left((1 + 2\mu_1\mu_2) \left(1 + x\eta \right) \left(1 + \frac{\eta}{x} \right) \right)^{-1}}{\left((1 + 2\mu_1\mu_2) \left(K_{1,S+1}\eta + K_{2,S+1}\eta \right) \right)^{m-1}} \right) (R_S - 1) \right] + 1
\end{aligned}$$

Recall that $K_{1,S+1} := \sum_{i=1}^m \frac{1}{u_{2i}^{S+1}}$ and $K_{2,S+1} := \sum_{i=1}^m u_{2i}^{S+1}$, and note that by (4.56)

and (4.59) we get $\mathbb{E}[K_{i,S+1} | \mathcal{F}_S] \leq \eta$ for $i = 1, 2$. Hence

$$\begin{aligned}
& \mathbb{E} \left[\left(1 - \frac{1}{4} \frac{\left((1 + 2\mu_1\mu_2) (1 + x\eta) \left(1 + \frac{\eta}{x} \right) \right)^{-1}}{\left((1 + 2\mu_1\mu_2) (K_{1,S+1}\eta + K_{2,S+1}\eta) \right)^{m-1}} \right) (R_S - 1) \right] + 1 \\
& \leq \mathbb{E} \left[\mathbb{E} \left[\left(1 - \frac{1}{4} \frac{\left((1 + 2\mu_1\mu_2) (1 + x\eta) \left(1 + \frac{\eta}{x} \right) \right)^{-1}}{\left((1 + 2\mu_1\mu_2) (K_{1,S+1}\eta + K_{2,S+1}\eta) \right)^{m-1}} \right) | \mathcal{F}_S \right] (R_S - 1) \right] + 1 \\
& \leq \left(1 - \frac{1}{4} \frac{\left((1 + 2\mu_1\mu_2) (1 + x\eta) \left(1 + \frac{\eta}{x} \right) \right)^{-1}}{\left((1 + 2\mu_1\mu_2) \left(\mathbb{E}[K_{1,S+1} | \mathcal{F}_S] \eta + \mathbb{E}[K_{2,S+1} | \mathcal{F}_S] \eta \right) \right)^{m-1}} \right) \mathbb{E}[(R_S - 1)] + 1 \\
& \leq r \mathbb{E}[(R_S - 1)] + 1 \tag{4.61}
\end{aligned}$$

where $r := 1 - \frac{1}{4} \left(\frac{1}{2\eta^2(1+2\mu_1\mu_2)} \right)^{m-1} \left(\frac{1}{(1+2\mu_1\mu_2)(1+x\eta)(1+\frac{\eta}{x})} \right)$. The transition to the last line of (4.61) is justified by Jensen's inequality. Also note that unlike in (4.19) where we concluded $\mathbb{E}[R_{S+2}] \leq r \mathbb{E}[(R_S - 1)] + 1$, we have $\mathbb{E}[R_{S+1}] \leq r \mathbb{E}[(R_S - 1)] + 1$. This is a result of directly using $\{K_{1,S+1}, K_{2,S+1}\}$ without having to resort to a process like D_{S+1} , and implies that the factor $\frac{1}{2}$ in the exponent of r in (4.19) can now be omitted. Therefore if we set $J_t = K_{1,t} + K_{2,t}$ and define d' as before, then by the results of Lemma 4.12 and Lemma 4.13 and inequality (4.28), we get

$$\mathbb{E}[R_{t+1} | J_0 \leq \eta] \leq 1 + 3r \lfloor \frac{t}{d'} \rfloor (R_0 - 1) \tag{4.62}$$

Proof of Theorem 4.4. The proof is identical to that of Theorem 4.2. \square

We can confirm that for every n the condition $\max\{\zeta_2, \dots, \zeta_{2m}, \xi_2, \dots, \xi_{2m}\} < 1$

is not vacuous, hence the previous results are applicable for certain parameter values. Observe first that if we set $a_i := a$ for all i , then

$\zeta_2 = \frac{4a}{4(2a-1)} \left(1 + \sum_{j=2}^m \prod_{k=2}^j \frac{2a}{4(2a-1)} \right) < \frac{4a}{4(2a-1)} \cdot \frac{4(2a-1)}{2(3a-2)}$, which is less than 1 whenever $a > 2$. Similarly, we conclude that whenever $a > 5$, ζ_{2m}, ξ_2 and ξ_{2m} are all less than 1. Now for $2 \leq i \leq m-1$, we get

$$\begin{aligned} \zeta_{2i} &= \frac{a}{2(2a-1)} & (4.63) \\ &+ \left(\frac{a}{(2a-1)} + \frac{a}{2(2a-1)} \frac{a}{2(2a-1)} \right) \left(1 + \sum_{j=i+1}^m \prod_{k=i+1}^j \frac{a}{2(2a-1)} \right) \\ &\leq q + (2q + q^2) \left(\frac{1 - q^{m+1}}{1 - q} \right) \end{aligned}$$

where $q = \frac{a}{2(2a-1)}$. Since $q \rightarrow \frac{1}{4}$ as $a \rightarrow \infty$, we conclude that $\zeta_{2i} \rightarrow \frac{1}{4} + \frac{3}{4} \left(1 - \left(\frac{1}{4} \right)^{m+1} \right) < 1$. Hence for a large enough, $\zeta_{2i} < 1$ for all j , and a similar deduction follows for ξ_{2i} .

The methods we used in this section used properties that were specific to the Gibbs samplers that we worked with. Some may be adaptable to other Gibbs samplers and other similar problems. However, at present it is not clear to us how to generalize the main results given in this section, for a more general class of Gibbs samplers.

4.12 Simulated results

We can attempt to assess the efficiency of some of our theoretically derived bounds by considering simulations and what rate of convergence they may suggest.

Our first look will be at the iterated function system discussed in Section 2.1 (recall that the system is based on the random functions $f_t(v) = \gamma_2^t / \left(\frac{\gamma_1^t}{x+v} + \frac{\gamma_3^t}{b+v} \right)$), in particular we will take a look at the same example considered at the end of the section - with $x = 2$, $b = 1$ and $a_i = i + 1$ for $i = 1, 2, 3, 4$. Observe that (also shown with more generality in [5]) if $\mathbb{E} \left[|F_{t+1}(v) - F_t(v)| \right] \leq r^t C_v$ for some $r < 1$ and C_v , then $\mathbb{E} \left[|F_\infty(v) - F_t(v)| \right] \leq r^t \frac{C_v}{1-r}$. If this is indeed the case, then we can possibly gain insight into the magnitude of r by approximating the function $\ln \left(\mathbb{E} \left[|F_{t+1}(v) - F_t(v)| \right] \right)$.

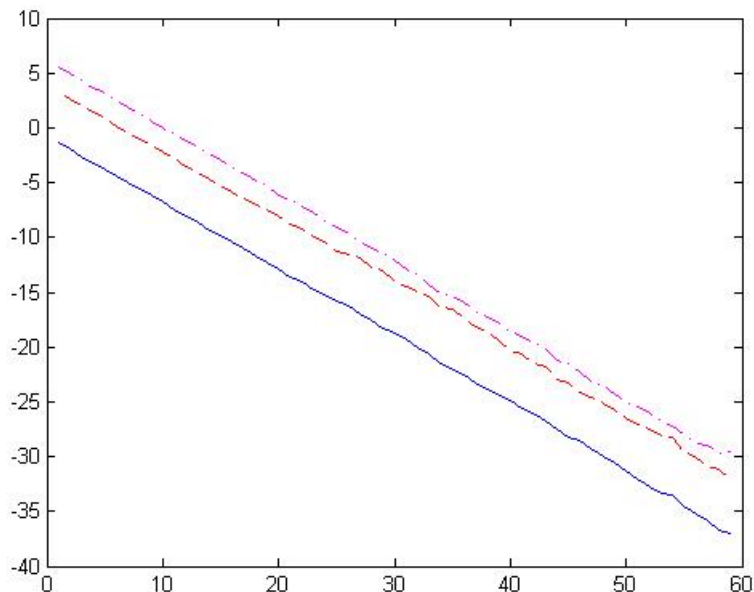
We do this by taking 1000 instances $\left\{ \left\{ a_{v,i}(t) \right\}_{i=1}^{1000} \right\}_{t=1}^{60}$ over 60 time steps of the backwards iterated function system corresponding to this problem, for each of the initial values $v = 1, 100$ and 1000 . Thus for each $1 \leq i \leq 1000$, $a_{v,i}(t)$ represents an independent simulation of the process $\{F_t(v)\}$ for $t = 1, \dots, 60$. For every $t = 1, \dots, 59$ we can take the approximation

$$\mathbb{E} \left[|F_{t+1}(v) - F_t(v)| \right] \approx \frac{1}{1000} \sum_{i=1}^{1000} |a_{v,i}(t+1) - a_{v,i}(t)|$$

Figure 4.1 is a plot of the function $\ln \left(\frac{1}{1000} \sum_{i=1}^{1000} |a_{v,i}(t+1) - a_{v,i}(t)| \right)$, with the solid line representing $v = 1$, while the dashed and dot-dashed lines represent

$v = 100$ and $v = 1000$ respectively. This does in fact support the relation

Figure 4.1: Simulated convergence for $n = 3$ case

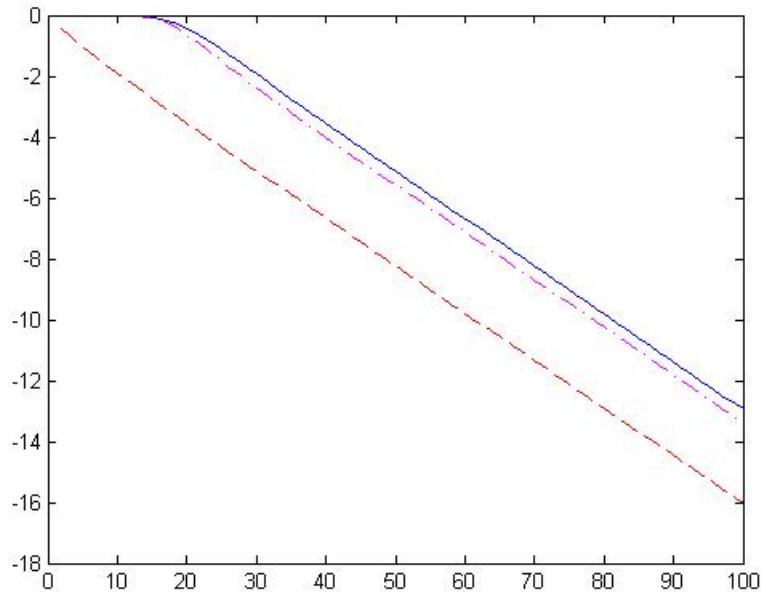


$\mathbb{E} \left[\left| F_{t+1}(v) - F_t(v) \right| \right] \leq r^t C_v$, and the slope suggests that $r \approx 0.55$. In particular, for $v = 1$ the simulations imply that we would only need to run the chain for 30 steps in order to obtain an independent sample that is (in the Wasserstein metric) at most 10^{-5} away from equilibrium.

We can also look at the process R_t defined in Section 4.4 and attempt to estimate through simulations how quickly it converges to 1. We consider the example in Section 4.8, with $x = 2$, $b = 3$ and $a_i = i$ for $i = 1, \dots, 5$. Figure 4.2 shows the mean of $\ln(R_t - 1)$ obtained from 10000 independent simulations of R_t , and taken over a period of 100 steps. The solid line represents initial conditions $U = (10^6, 10^6)$

and $V = 2U$, while the dashed and dot-dashed lines represent $U = (1, 1)$, $V = 2U$ and $U = (10^{-6}, 10^{-6})$, $V = 2U$. Observe that in all three cases $R_0 = 2$, yet there appears to be a clear delay in the convergence of the solid and dot-dashed lines. Our interpretation of this is that it relates to the development of our 'auxiliary processes' in Section 4.6. Recall that in our proof, geometric convergence of R_t required for U_0 and $\frac{1}{U_0}$ to be bounded from above by some constant η . The delay would thus be the time taken for U_t to satisfy these bounds.

Figure 4.2: Simulated convergence for $n = 4$ case



In addition, the slope of the lines in 4.2 suggests a value of $r \approx 0.86$, where r is defined in the paragraph following equation (4.19). For the example considered in Section 4.8, this would imply a significantly lower mixing time (no more than 50

iterations for this example) than the upper bound of 50,000,000 obtained in that section.

5 An Image Restoration Model

5.1 Introduction

A.L. Gibbs [3] introduced a stochastic image restoration model for an N pixel greyscale image $x = \{x_i\}_{i=1}^N$. More specifically, in this model each pixel x_i corresponds to a real value in $[0, 1]$, where a black pixel is represented by 0 and a white pixel is represented by the value 1. It is assumed that in the real-world space of such images, each pixel tends to be like its nearest neighbours (in the absence of any evidence otherwise). This assumption is expressed in the prior probability density of the image, which is given by

$$\pi_\gamma(x) \propto \exp \left\{ - \sum_{\langle i,j \rangle} \frac{1}{2} [\gamma (x_i - x_j)]^2 \right\} \quad (5.1)$$

on the state space $[0, 1]^N$, and is equal to 0 elsewhere. The sum in (5.1) is over all pairs of pixels that are considered to be neighbours, and the parameter γ represents the strength of the assumption that neighbouring pixels are similar. Here images are assumed to have an underlying graph structure. The familiar 2-dimensional

digital image is a special case, where usually one might assume that the neighbours of a pixel x_i in the interior of the image (i.e. x_i not on the boundary of the image) are the 4 or 8 pixels surrounding x_i , depending on whether or not we decide to consider the 4 pixels diagonal to x_i .

The actual observed image $y = \{y_i\}_{i=1}^N$ is assumed to be the result of the original image subject to distortion by random noise, with every pixel modified independently through the addition of a *Normal* $(0, \sigma^2)$ random variable (hence $y_i \in \mathbb{R}$). The resulting posterior probability density for the original image is given by

$$\pi_{\text{posterior}}(x|y) \propto \exp \left\{ - \sum_{i=1}^N \frac{1}{2\sigma^2} (x_i - y_i)^2 - \sum_{\langle i,j \rangle} \frac{1}{2} \left[\gamma (x_i - x_j) \right]^2 \right\} \quad (5.2)$$

supported on $[0, 1]$.

Samples from (5.2) can be approximately obtained by means of a Gibbs sampler. In this instance, the algorithm works as follows: at every iteration the sampler chooses a site i uniformly at random, and replaces the value x_i at this location according to the full conditional density at that site. This density is given by

$$\pi_{FC} \left(x_i | y, x_{k \neq i} \right) \propto \exp \left\{ \frac{(\sigma^{-2} + n_i \gamma^2)}{2} \cdot \left[x_i - (\sigma^{-2} + n_i \gamma^2)^{-1} \left(\sigma^{-2} y_i + \gamma^2 \sum_{j \sim i} x_j \right) \right]^2 \right\} \quad (5.3)$$

on $[0, 1]$ and 0 elsewhere. Here n_i is the number of neighbours the i^{th} pixel has, and

$j \sim i$ indicates that the j^{th} pixel is one of them. It follows that (5.3) is a restriction of a

$Normal \left((\sigma^{-2} + n_i \gamma^2)^{-1} \left(\sigma^{-2} y_i + \gamma^2 \sum_{j \sim i} x_j \right), (\sigma^{-2} + n_i \gamma^2)^{-1} \right)$ distribution to the set $[0, 1]$.

The bound on the rate of convergence to equilibrium given in [3] is stated in terms of the Wasserstein metric d_W . This is defined as follows: if μ_1 and μ_2 are two probability measures on the same state space which is endowed with some metric d , then

$$d_W(\mu_1, \mu_2) := \inf \mathbb{E} [d(\xi_1, \xi_2)]$$

where the infimum is taken over all joint distributions (ξ_1, ξ_2) such that $\xi_1 \sim \mu_1$ and $\xi_2 \sim \mu_2$. We will also use the convention $d_W(\xi_1, \xi_2) := d_W(\mu_1, \mu_2)$. Gibbs [3] shows that

Theorem 5.1. [1] *Let X^t be a copy of the Markov chain evolving according to the Gibbs sampler, and let Z^t be a chain in equilibrium, distributed according to $\pi_{\text{posterior}}$. Then $d_W(X^t, Z^t) \leq \epsilon$ whenever*

$$t > \vartheta(\epsilon) := \frac{\log \left(\frac{\epsilon}{n_{\max} N} \right)}{\log \left(1 - N^{-1} (1 + n_{\max} \gamma^2 \sigma^2)^{-1} \right)} \quad (5.4)$$

Here $n_{\max} := \max_i \{n_i\}$, and the underlying metric on the state space is defined as $d(x, z) := \sum_i n_i |x_i - z_i|$. This is a non-standard choice for a metric on $[0, 1]^N$, however it is comparable to the more usual l_1 taxicab metric $\hat{d}(x, y) := \sum_i |x_i - z_i|$

since

$$n_{min} \cdot \hat{d}(x, y) \leq d(x, y) \leq n_{max} \cdot \hat{d}(x, y)$$

where $n_{max} := \max_i \{n_i\}$ and $n_{min} := \min_i \{n_i\}$. Hence, for two probability measures μ_1 and μ_2 on $[0, 1]^N$, it follows immediately that

$$n_{min} \cdot d_{\hat{W}}(\mu_1, \mu_2) \leq d_W(\mu_1, \mu_2) \leq n_{max} \cdot d_{\hat{W}}(\mu_1, \mu_2)$$

where $d_{\hat{W}}$ and d_W are the Wasserstein metrics associated with \hat{d} and d respectively.

Remark 5.1. Equation (5.4) appears in [3] with the denominator being $\log \left(N - 1/N + n_{max} N^{-1} \gamma^2 (\sigma^{-2} + n_{max} \gamma^2)^{-1} \right)$. It is obvious from their proof that this is a typographical error, and that the term $N - 1/N$ was intended to be $(N - 1) / N$.

The most commonly used metric for measuring the distance of a Markov chain from its equilibrium distribution is the total variation metric, defined for two probability measures μ_1 and μ_2 on Ω by

$$d_{TV}(\mu_1, \mu_2) := \sup |\mu_1(A) - \mu_2(A)|$$

where the supremum is taken over all measurable $A \subseteq \Omega$. For two random variables ξ_1 and ξ_2 , we define $d_{TV}(\xi_1, \xi_2)$ to be $d_{TV}(\mu_1, \mu_2)$, where $\xi_i \sim \mu_i$.

It is not difficult to see that d_{TV} is a special case of d_W when the underlying metric is given by $d(x, z) = 1$ if $x \neq z$. In general however, convergence in d_W does not imply convergence in d_{TV} , and vice versa (see [4] for examples where

convergence fails, as well as some conditions under which convergence in one of d_W , d_{TV} implies convergence in the other). The purpose of this paper is to obtain a bound in d_{TV} by making use of (5.4) and simple properties of the Markov chain, without specifically engaging in a new study of the mixing time.

Let X_t be a copy of the Markov chain, and let μ^t be its probability distribution.

Furthermore, define

$$\zeta_i := (\sigma^{-2} + n_i \gamma^2)^{-1} (\sigma^{-2} y_i + \gamma^2 n_{max}), \quad \zeta := \max \{|\zeta_i|\} \quad \text{and} \quad \tilde{\sigma}_i^2 = (\sigma^{-2} + n_i \gamma^2)^{-1}.$$

If π is the posterior distribution with density function $\pi_{posterior}$, we show that

Theorem 5.2. *Let X_t be a copy of the Markov chain evolving according to the Gibbs sampler, and let Z^t be a chain in equilibrium. Then $d_{TV}(X^t, Z^t) \leq \epsilon$ whenever*

$$t > \vartheta(\omega^2) + M \tag{5.5}$$

where $M = \left\lceil N \log(N) + N \log\left(\frac{2}{\epsilon}\right) \right\rceil$ and $\omega = \left[1 - \left(1 - \frac{\epsilon}{2}\right)^{M-1} \right] / \left(1 + e^{\frac{(\zeta+1)^2}{2\tilde{\sigma}^2}} \right)$.

Akin to the bound for the metric d_W , this bound is also $O(N \log \frac{N}{\epsilon})$. A notable difference, however, is that in our bound there is a (quadratic) dependence on ζ (and hence a quadratic dependence on $\max \{|y_i|\}$).

5.2 From d_W to d_{TV}

Let t be some fixed time, and let X^s and Z^s ($s = 1, \dots, t$) be two instances of the Markov chain, evolving as defined in the lines preceding (5.3). The coupling

method [2] allows us to bound total variation via the inequality

$$d_{TV}(X^t, Z^t) \leq \mathbb{P}[X^t \neq Z^t].$$

Having uniformly selected i from $\{1, \dots, N\}$, we couple the pixel X_i^{t+1} with Z_i^{t+1} as follows: let f_i and g_i be the conditional density functions of X_i^{t+1} given X^t and of Z_i^{t+1} given Z^t , respectively. Choose a point (a_1, a_2) uniformly from the area defined by $A_X = \{(a, b) \mid f_i(a) > 0, 0 \leq b \leq f_i(a)\}$ - i.e. the area under the graph of f_i , and set $X_i^{t+1} = a_1$. If the point (a_1, a_2) is also in the set $A_Z = \{(a, b) \mid g_i(a) > 0, 0 \leq b \leq g_i(a)\}$, then set $Z_i^{t+1} = X_i^{t+1} = a_1$. Otherwise $(a_1, a_2) \in A_X \setminus A_Z$, and in this case choose a point (b_1, b_2) uniformly from $A_Z \setminus A_X = \{(a, b) \mid g_i(a) \geq b \geq f_i(a)\}$ and set $Z_i^{t+1} = b_1$. Observe that X^s and Z^s ($s = 0, \dots, t+1$) are indeed two faithful copies of the Markov chain.

In order to proceed, we will establish the following results.

Lemma 5.3. *Let $U_1 \sim \text{Normal}(\mu_1, \sigma^2)$ and $U_2 \sim \text{Normal}(\mu_2, \sigma^2)$, and let W_1 and W_2 have the distributions of U_1 and U_2 conditioned to be in some measurable set S . Let f_{U_1} , f_{U_2} , f_{W_1} and f_{W_2} be their respective density functions. Then*

$$d_{TV}(W_1, W_2) \leq \frac{d_{TV}(U_1, U_2)}{\min(\int_S f_{U_1}, \int_S f_{U_2})}$$

Proof. We start by noting that

$$\begin{aligned}
d_{TV}(W_1, W_2) &= \int_{f_{W_1} \geq f_{W_2}} (f_{W_1} - f_{W_2}) \\
&= \int_{f_{W_1} \geq f_{W_2}} \left(\frac{f_{U_1}}{\int_S f_{U_1}} - \frac{f_{U_2}}{\int_S f_{U_2}} \right)
\end{aligned} \tag{5.6}$$

The first equality is one of a few different equivalent definitions of total variation.

A proof is given in Proposition 3 of [1].

Now if $\int_S f_{U_1} \geq \int_S f_{U_2}$, then the above is bounded by

$$\begin{aligned}
d_{TV}(W_1, W_2) &\leq \frac{1}{\int_S f_{U_2}} \int_{f_{W_1} \geq f_{W_2}} (f_{U_1} - f_{U_2}) \\
&\leq \frac{1}{\int_S f_{U_2}} \int_{f_{U_1} \geq f_{U_2}} (f_{U_1} - f_{U_2}) \\
&= \frac{d_{TV}(U_1, U_2)}{\min(\int_S f_{U_1}, \int_S f_{U_2})}
\end{aligned} \tag{5.7}$$

The second inequality follows from the observation that

$$\frac{f_{U_1}(w)}{\int_S f_{U_1}} \geq \frac{f_{U_2}(w)}{\int_S f_{U_2}} \Rightarrow \frac{f_{U_1}(w)}{\int_S f_{U_2}} \geq \frac{f_{U_2}(w)}{\int_S f_{U_2}} \Rightarrow f_{U_1}(w) \geq f_{U_2}(w)$$

Similarly, if $\int_S f_{U_2} \geq \int_S f_{U_1}$, then we repeat the same argument with

$$d_{TV}(W_1, W_2) = \int_{f_{W_2} \geq f_{W_1}} (f_{W_2} - f_{W_1})$$

in place of (5.6), arriving at the same result. \square

A simple but useful result is the following lemma:

Lemma 5.4. $(2\pi\sigma^2)^{-1/2} \int_0^1 e^{-\frac{(x-\zeta_i)^2}{2\sigma^2}} \geq (2\pi\sigma^2)^{-1/2} e^{-\frac{(|\zeta_i|+1)^2}{2\sigma^2}}$

Proof. This is trivial, since $(|\zeta_i| + 1) \geq |x - \zeta_i|$ for any $x \in [0, 1]$. \square

Now let $U_1 \sim \text{Normal} \left((\sigma^{-2} + n_i \gamma^2)^{-1} (\sigma^{-2} y_i + \gamma^2 \sum_{j \sim i} x_j^t), \tilde{\sigma}_i^2 \right)$ and $U_2 \sim \text{Normal} \left((\sigma^{-2} + n_i \gamma^2)^{-1} (\sigma^{-2} y_i + \gamma^2 \sum_{j \sim i} z_j^t), \tilde{\sigma}_i^2 \right)$. Applying Lemma 5.3 to (X_i^{t+1}, Z_i^{t+1}) with $S = [0, 1]$, we see that conditional on \mathcal{F}_t (sigma algebra generated by X^t and Z^t)

$$\begin{aligned} \mathbb{P} [X_i^{t+1} \neq Z_i^{t+1} | \mathcal{F}_t] &= d_{TV} (X_i^{t+1}, Z_i^{t+1} | \mathcal{F}_t) \\ &\leq \frac{d_{TV} (U_1, U_2 | \mathcal{F}_t)}{\min (\int_S f_{U_1}, \int_S f_{U_2})} \\ &\leq (2\pi \tilde{\sigma}_i^2)^{1/2} e^{\frac{(|\zeta_i|+1)^2}{2\tilde{\sigma}_i^2}} d_{TV} (U_1, U_2 | \mathcal{F}_t) \end{aligned} \quad (5.8)$$

For the second inequality we have used Lemma 5.4. By Lemma 15 of [4] it follows that

$$d_{TV} (U_1, U_2 | \mathcal{F}_t) \leq \frac{|\mathbb{E} [U_1 | \mathcal{F}_t] - \mathbb{E} [U_2 | \mathcal{F}_t]|}{\sqrt{2\pi \tilde{\sigma}_i^2}} \quad (5.9)$$

Hence by (5.8)

$$\begin{aligned} \mathbb{P} [X_i^{t+1} \neq Z_i^{t+1} | \mathcal{F}_t] &\leq e^{\frac{(|\zeta_i|+1)^2}{2\tilde{\sigma}_i^2}} \left| \mathbb{E} [U_1 | \mathcal{F}_t] - \mathbb{E} [U_2 | \mathcal{F}_t] \right| \\ &= e^{\frac{(|\zeta_i|+1)^2}{2\tilde{\sigma}_i^2}} \tilde{\sigma}_i^2 \gamma^2 \left| \sum_{j \sim i} X_j^t - \sum_{j \sim i} Z_j^t \right| \\ &\leq e^{\frac{(|\zeta_i|+1)^2}{2\tilde{\sigma}_i^2}} \tilde{\sigma}_i^2 \gamma^2 \sum_{j \sim i} |X_j^t - Z_j^t| \end{aligned} \quad (5.10)$$

We can now proceed with the proof of Theorem 5.2.

Proof of Theorem 5.2. Let $\epsilon > 0$ be given, and define $\tilde{\epsilon} := 1 - \left(1 - \frac{\epsilon}{2}\right)^{M^{-1}}$ (recall that

$M = \left\lceil N \log(N) + N \log\left(\frac{2}{\epsilon}\right) \right\rceil$) and $\omega := \tilde{\epsilon} / \left(1 + e^{\frac{(\zeta+1)^2}{2\tilde{\sigma}^2}}\right)$ with $\tilde{\sigma} := \min\{\tilde{\sigma}_i\}$. By

Theorem 5.1, $d_W(X^t, Z^t) \leq \omega^2$ whenever

$t \geq \tau := \left\lceil \log\left(\frac{\omega^2}{n_{max}N}\right) / \log\left(1 - N^{-1}(1 + \sigma^2 n_{max} \gamma^2)^{-1}\right) \right\rceil$. Since the infimum in

the definition of d_W is achieved (see for example Section 5.1 of [11]), we can find

a joint distribution $\mathcal{L}(u^\tau, v^\tau)$ of two random variables $u^\tau \sim X^\tau$ and $v^\tau \sim Z^\tau$, such

that $\mathbb{E}[d(u^\tau, v^\tau)] = \mathbb{E}[\sum n_i |u_i^\tau - v_i^\tau|] \leq \omega^2$ (we use the superscript τ in u^τ and v^τ

to preserve notational consistency with X^τ and Z^τ). And by Markov's inequality

we get

$$\begin{aligned} \mathbb{P}\left[\sum_{k \sim j} |u_k^\tau - v_k^\tau| \geq \omega \text{ for some } j\right] &\leq \mathbb{P}[d(u^\tau, v^\tau) \geq \omega] \\ &\leq \omega \end{aligned} \tag{5.11}$$

For $s = 1, \dots$, define the Markov chains $u^{\tau+s} \sim X^{\tau+s}$ and $v^{\tau+s} \sim Z^{\tau+s}$ by uniformly

choosing (for every s) a site i and assigning values to $(u_i^{\tau+s}, v_i^{\tau+s})$ as described at

the beginning of Section 5.2. Note that $d_{TV}(u^{\tau+s}, v^{\tau+s}) = d_{TV}(X^{\tau+s}, Z^{\tau+s})$, hence

it suffices to show that $d_{TV}(u^{\tau+s}, v^{\tau+s}) \leq \epsilon$ whenever $t > \vartheta(\omega^2) + M$. By splitting

up the above probability and applying (5.10) and (5.11), we conclude that at the

chosen site i

$$\begin{aligned}
\mathbb{P} [u_i^{\tau+1} \neq v_i^{\tau+1}] &= \mathbb{P} \left[u_i^{\tau+1} \neq v_i^{\tau+1} \left| \sum_{k \sim i} |u_k^\tau - v_k^\tau| < \omega \right. \right] \cdot \mathbb{P} \left[\sum_{k \sim i} |u_k^\tau - v_k^\tau| < \omega \right] \\
&\quad + \mathbb{P} \left[u_i^{\tau+1} \neq v_i^{\tau+1} \left| \sum_{k \sim i} |u_k^\tau - v_k^\tau| \geq \omega \right. \right] \cdot \mathbb{P} \left[\sum_{k \sim i} |u_k^\tau - v_k^\tau| \geq \omega \right] \\
&\leq e^{\frac{(|\zeta_i|+1)^2}{2\tilde{\sigma}_i^2}} \tilde{\sigma}_i^2 \gamma^2 \omega + \omega \\
&\leq \omega \left(e^{\frac{(\zeta+1)^2}{2\tilde{\sigma}^2}} + 1 \right) \\
&= \tilde{\epsilon}
\end{aligned} \tag{5.12}$$

Let i_m be the pixel chosen at time $\tau + m$ for $m = 1, 2, \dots$. For $j \geq 1$, define the events $B_j := \{u_{i_j}^{\tau+j} = v_{i_j}^{\tau+j}\}$ and $B_0 := \{d(u^\tau, v^\tau) \leq \omega\}$, and observe that in the event $\{\bigcap_{k=0}^j B_k\}$, we have $d(u^{\tau+j}, v^{\tau+j}) \leq d(u^\tau, v^\tau) \leq \omega$. Therefore by equations (5.10) and (5.11)

$$\begin{aligned}
\mathbb{P} \left[u_{i_m}^{\tau+m} \neq v_{i_m}^{\tau+m} \left| \bigcap_{k=1}^{m-1} B_k \right. \right] &\leq \mathbb{P} \left[u_{i_m}^{\tau+m} \neq v_{i_m}^{\tau+m} \left| \bigcap_{k=0}^{m-1} B_k \right. \right] \mathbb{P} [B_0] + \omega \\
&\leq \omega \left(e^{\frac{(\zeta+1)^2}{2\tilde{\sigma}^2}} + 1 \right) \\
&= \tilde{\epsilon}
\end{aligned}$$

By induction on m we get that

$$\begin{aligned}
\mathbb{P} \left[\bigcap_{j=1}^m B_j \right] &\geq \mathbb{P} \left[B_m \left| \bigcap_{j=1}^{m-1} B_j \right. \right] \cdot \mathbb{P} \left[\bigcap_{j=1}^{m-1} B_j \right] \\
&\geq (1 - \tilde{\epsilon})^m
\end{aligned} \tag{5.13}$$

Note that the case $m = 1$ follows directly from (5.12). We will now refer to the 'coupon collector' problem, discussed in section 2.2 of [7]: if θ is the first time when a coupon collector has obtained all N out of N coupons, then

$$\mathbb{P}[\theta > M] \leq \frac{\epsilon}{2} \quad (5.14)$$

Let $\phi := \tau + M$ and let $\theta := \min \{l \geq 1 : \{1, \dots, N\} \subseteq \{i_1, \dots, i_l\}\}$ - i.e. $\tau + \theta$ is the first time when every pixel site has been chosen at least once after τ . Then

$$\begin{aligned} \mathbb{P}[u^\phi \neq v^\phi] &= \mathbb{P}[u^\phi \neq v^\phi | \theta > M] \cdot \mathbb{P}[\theta > M] + \mathbb{P}[u^\phi \neq v^\phi, \theta \leq M] \\ &\leq \mathbb{P}[\theta > M] + \mathbb{P}[u_{i_j}^{\tau+j} \neq v_{i_j}^{\tau+j} \text{ for some } 1 \leq j \leq M] \\ &= \mathbb{P}[\theta > M] + 1 - \mathbb{P}\left[\bigcap_{j=1}^M B_j\right] \\ &\leq \frac{\epsilon}{2} + 1 - (1 - \tilde{\epsilon})^M \\ &\leq \epsilon \end{aligned} \quad (5.15)$$

This proves the statement of the theorem. □

Bibliography

- [1] G.O. Roberts and J.S. Rosenthal (2004), General state space Markov chains and MCMC algorithms. *Probability Surveys*, Vol. 1, 20-71
- [2] T. Lindvall (1992), *Lectures on the Coupling Method*. Wiley&Sons, New York.
- [3] A.L. Gibbs (2004), Convergence in the Wasserstein Metric for Markov Chain Monte Carlo Algorithms with Application to Image Restoration, *Stochastic Models*, Vol 20, Issue 4, 473-492
- [4] N. Madras and D. Sezer (2010), Quantitative bounds for Markov chain convergence: Wasserstein and total variation distances, *Bernoulli*, Vol. 16, No. 3, 882-908
- [5] D. Steinsaltz (1999), Locally contractive iterated function systems, *Annals of Probability*, Vol. 27, No. 4, 1952-1979
- [6] N. Madras and H. Liu (2010), Random pattern-avoiding permutations, *Contemporary Mathematics* vol. 520, 173-194
- [7] D.A. Levin and Y. Peres and E. L. Wilmer (2008), *Markov Chains and Mixing Times*, AMS, Providence
- [8] R. Martin and D. Randall (2006), Disjoint decomposition of Markov chains and sampling circuits in Cayley graphs, *Combinatorics Probability and Computing*, Vol. 15, 411-448
- [9] J. Chen and H. Rubin (1986), Bounds for the difference between median and mean of gamma and Poisson distributions, *Statistics & Probability Letters* 4, 281-283
- [10] G.O. Roberts and J.S. Rosenthal (2002), One-shot coupling for certain stochastic recursive sequences, *Stochastic Processes and their Applications*, Vol. 99, 195-208

- [11] M.F.Chen (1992), From Markov Chains to Non-Equilibrium Particle Systems, World Scientific, Singapore
- [12] S. Geman, D. Geman (1984), Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, IEEE Trans. on pattern analysis and machine intelligence Vol. 6, 721-741
- [13] W.R. Gilks, S. Richardson (1996), D.J. Spiegelhalter, Markov chain Monte Carlo in practice, Chapman and Hall, London
- [14] A.F.M. Smith, G.O. Roberts (1993), Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods, Journal of the Royal Statistical Society: Series B, Vol. 55, 3-24
- [15] L. Tierney (1994), Markov chains for exploring posterior distributions, Annals of Statistics, Vol. 22, 1701-1762
- [16] O. Jovanovski, N. Madras (2013), Convergence Rates for Hierarchical Gibbs Samplers, submitted
- [17] O. Jovanovski (2013), A bound in total variation for a stochastic image restoration model, submitted
- [18] D. Revuz (2008), Markov Chains, North Holland, Amsterdam
- [19] N. Madras (2002), Lectures on Monte Carlo Methods, Fields Institute Monographs - AMS, Providence
- [20] N. Madras, D. Randall (2002), Markov Chain Decomposition for Convergence Rate Analysis, The Annals of Applied Probability, Vol. 12, No. 2, 581-606
- [21] R. Martin, D. Randall (2006), Disjoint decomposition of Markov chains and sampling circuits in Cayley graphs, Combinatorics, Probability and Computing Vol. 15, 411-448
- [22] M. Dyer, C. Greenhill (1998), A more rapidly mixing Markov chain for graph colorings, Random Structures and Algorithms Vol. 13, 285-317
- [23] P. Diaconis, L. Saloff-Coste (1993), Comparison theorems for reversible Markov chains, Annals of Applied Probability Vol. 3, 696-730
- [24] A. Zellner, C.K. Min (1995), Gibbs sampler convergence criteria, Journal of the American Statistical Association, Vol. 90, Issue 431, 921-927

- [25] D.A. van Dyk, T. Park (2008), Partially Collapsed Gibbs Samplers, *Journal of the American Statistical Association*, Vol. 103, Issue 482, 283-305
- [26] V. Witkovsky (2001), Computing the distribution of a linear combination of inverted gamma variables, *Kybernetika*, Vol. 37, Issue 1, 79-90
- [27] J. Basag, P. Green, D. Higdon, K. Mengersen (1995), Bayesian Computation and Stochastic Systems, *Statistical Science*, Vol. 10, No. 1, 3-66
- [28] P. Cowans (2004), Information retrieval using hierarchical Dirichlet process, *Proceedings of the Annual International Conference on Research and Development in Information Retrieval*, Vol. 27, pp. 564-565
- [29] M. Stephens, N. Smith and P. Donnelly (2001), A new statistical method for haplotype reconstruction from population data, *American Journal of Human Genetics*, Vol. 68, pp. 978–989.
- [30] P. E. Rossi, G. M. Allenby and R. McCulloch (2006), Bayesian Statistics and Marketing, *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, Vol. 169, Issue 4.
- [31] F.J. Dez, J. Mira, E. Iturralde and S. Zubillaga (1997), DIAVAL a Bayesian expert system for echocardiography, *Artificial Intelligence in Medicine*, Vol. 10 , pp. 59–73.
- [32] Jiang X, Cooper GF (2010), A Bayesian spatio-temporal method for disease outbreak detection, *J Am Med Inform Assoc*, Vol. 17, pp. 462–71.
- [33] A. Constantinou, N. Fenton, M. Neil (2012). pi-football: A Bayesian network model for forecasting Association Football match outcomes, *Knowledge-Based Systems*, Vol. 36, pp. 322–339.
- [34] X. Jiang, R. Neapolitan, M. Barmada, S. Visweswaran (2011), Learning Genetic Epistasis using Bayesian Network Scoring Criteria, *BMC Bioinformatics*, 12: 89.

Appendix A

Proof of Lemma 4.16. Note that for the case $m = 2$, inequality (4.42) is true as this corresponds to $n = 4$ or $n = 5$. For $m = 3$, (4.40) amounts to

$$\begin{aligned} \frac{F_{(3,2)}^1(v)}{F_{(3,2)}^1(u)} &= \frac{f_{(3,2)}^1(v)}{f_{(3,2)}^1(u)} \\ &= \frac{v_2}{u_2} \left(\frac{\frac{\gamma_1}{\frac{x}{u_2}+1} + \frac{\gamma_3}{1+\frac{u_4}{u_2}}}{\frac{\gamma_1}{\frac{x}{v_2}+1} + \frac{\gamma_3}{1+\frac{v_4}{v_2}}} \right) \\ &< \frac{v_2}{u_2} \end{aligned}$$

and similarly

$$\begin{aligned} \frac{F_{(3,6)}^1(v)}{F_{(3,6)}^1(u)} &= \frac{v_6}{u_6} \left(\frac{\frac{\gamma_7}{\frac{b}{u_6}+1} + \frac{\gamma_5}{1+\frac{u_4}{u_6}}}{\frac{\gamma_7}{\frac{b}{v_6}+1} + \frac{\gamma_5}{1+\frac{v_4}{v_6}}} \right) \\ &< \frac{v_6}{u_6} = \frac{v_2}{u_2} \end{aligned}$$

while (4.41) follows from

$$\begin{aligned} \frac{F_{(3,4)}^1(v)}{F_{(3,4)}^1(u)} &= \frac{v_4}{u_4} \left(\frac{\frac{\gamma_3}{1+\frac{u_2}{u_4}} + \frac{\gamma_5}{1+\frac{u_6}{u_4}}}{\frac{\gamma_3}{1+\frac{v_2}{v_4}} + \frac{\gamma_5}{1+\frac{v_6}{v_4}}} \right) \\ &= \frac{v_4}{u_4} = \frac{v_2}{u_2} \end{aligned}$$

This in turn gives (4.42): for $j = 1$ this becomes

$$\begin{aligned}
\frac{F_{(3,2)}^{\lceil \frac{m}{2} \rceil}(v)}{F_{(3,2)}^{\lceil \frac{m}{2} \rceil}(u)} &= \left(\frac{\frac{\gamma_1}{F_{(3,2)}^1(u)+x} + \frac{\gamma_3}{F_{(3,2)}^1(u)+F_{(3,4)}^1(u)}}{\frac{\gamma_1}{F_{(3,2)}^1(v)+x} + \frac{\gamma_3}{F_{(3,2)}^1(v)+F_{(3,4)}^1(v)}} \right) \\
&= \frac{v_2}{u_2} \left(\frac{\frac{\frac{\gamma_1}{F_{(3,2)}^1(u)} + \frac{x}{u_2}}{\frac{\gamma_1}{F_{(3,2)}^1(v)} + \frac{x}{v_2}} + \frac{\frac{\gamma_3}{F_{(3,2)}^1(u)} + \frac{F_{(3,4)}^1(u)}{u_2}}{\frac{\gamma_3}{F_{(3,2)}^1(v)} + \frac{F_{(3,4)}^1(v)}{v_2}}}{\frac{\gamma_1}{F_{(3,2)}^1(v)} + \frac{\gamma_3}{F_{(3,2)}^1(v)+F_{(3,4)}^1(v)}} \right) \\
&< \frac{v_2}{u_2}
\end{aligned}$$

since $\frac{x}{u_2} > \frac{x}{v_2}$ and $\frac{F_{(3,2)}^1(u)}{u_2} > \frac{F_{(3,2)}^1(v)}{v_2}$ since we established (4.40), and $\frac{F_{(3,4)}^1(u)}{u_2} = \frac{F_{(3,4)}^1(v)}{v_2}$ by (4.41). Similar arguments follow for $j = 2, 3$. Suppose that the lemma

holds for $m = k$, where k is odd. Then for $j \in \left\{ 1, \dots, \left\lceil \frac{k+1}{2} \right\rceil \right\}$ we have

$$F_{(k+1,2j)}^{\lceil \frac{k+1}{2} \rceil - 1}(u_2, \dots, u_{2(k+1)}) = F_{(k,2j)}^{\lceil \frac{k}{2} \rceil - 1}(u_2, \dots, u_{2k})$$

since $F_{(k+1,2j)}^{\lceil \frac{k+1}{2} \rceil - 1}$ does not depend on $u_{2(k+1)}$ (note that $F_{(m,2l)}^c(u)$ does not depend on any coordinate to the right of u_{2l+2c} or to the left of u_{2l-2c}), and the continued fractions representing these two functions will be identical. For this reason, if $j \in \left\{ 1, \dots, \left\lceil \frac{k+1}{2} \right\rceil - 1 \right\}$ then (4.40) remains true by the inductive hypothesis, and for $j = \left\lceil \frac{k+1}{2} \right\rceil$ the equality in (4.41) is partially satisfied, in the sense that

$$\frac{F_{(k+1,2\lceil \frac{k+1}{2} \rceil)}^{\lceil \frac{k+1}{2} \rceil - 1}(v)}{F_{(k+1,2\lceil \frac{k+1}{2} \rceil)}^{\lceil \frac{k+1}{2} \rceil - 1}(u)} = \frac{F_{(k,2\lceil \frac{k}{2} \rceil)}^{\lceil \frac{k}{2} \rceil - 1}(v)}{F_{(k,2\lceil \frac{k}{2} \rceil)}^{\lceil \frac{k}{2} \rceil - 1}(u)} = \frac{v_2}{u_2}$$

Similarly, if we let $\tilde{F}_{(k,2j)}^{\lceil \frac{k}{2} \rceil - 1}$ be the function $F_{(k,2j)}^{\lceil \frac{k}{2} \rceil - 1}$ with $(\gamma_1^t, \gamma_2^t, \dots, \gamma_{2k+1}^t)$ replaced by $(\gamma_3^t, \gamma_4^t, \dots, \gamma_{2k+3}^t)$ for $t \in \left\{ 1, \dots, \left\lceil \frac{k}{2} \right\rceil - 1 \right\}$, then for $j \in \left\{ \left\lceil \frac{k+1}{2} \right\rceil + 1, \dots, k+1 \right\}$

we get by the same reasoning

$$F_{(k+1,2j)}^{\lceil \frac{k+1}{2} \rceil - 1}(u_2, \dots, u_{2(k+1)}) = \tilde{F}_{(k,2(j-1))}^{\lceil \frac{k}{2} \rceil - 1}(u_4, \dots, u_{2(k+1)})$$

The inductive hypothesis for $F_{(k,2j)}^{\lceil \frac{k}{2} \rceil - 1}$ extends $\tilde{F}_{(k,2j)}^{\lceil \frac{k}{2} \rceil - 1}$ since these two functions are structurally identical'. We can therefore conclude for $m = k + 1$ the validity of inequality (4.40) for $j \in \left\{ \left\lceil \frac{k+2}{2} \right\rceil + 1, \dots, k + 1 \right\}$, and equality (4.41) for $j = \left\lceil \frac{k+1}{2} \right\rceil + 1$.

Now we are in a position to verify (4.42) for $m = k + 1$: let $\kappa = \left\lceil \frac{k+1}{2} \right\rceil$ and set

$$F_{(k+1,0)}^\kappa \equiv x, F_{(k+1,k+2)}^\kappa \equiv b$$

$$\begin{aligned} \frac{F_{(k+1,2j)}^\kappa(v)}{F_{(k+1,2j)}^\kappa(u)} &= \frac{\frac{\gamma_{2j-1}}{F_{(k+1,2(j-1))}^{\kappa-1}(u) + F_{(k+1,2j)}^{\kappa-1}(u)} + \frac{\gamma_{2j+1}}{F_{(k+1,2j)}^{\kappa-1}(u) + F_{(k+1,2(j+1))}^{\kappa-1}(u)}}{\frac{\gamma_{2j-1}}{F_{(k+1,2(j-1))}^{\kappa-1}(v) + F_{(k+1,2)}^{\kappa-1}(v)} + \frac{\gamma_{2j+1}}{F_{(k+1,2j)}^{\kappa-1}(v) + F_{(k+1,2(j+1))}^{\kappa-1}(v)}} \\ &= \frac{v_{2j}}{u_{2j}} \cdot \left(\frac{\frac{\frac{\gamma_{2j-1}}{F_{(k+1,2(j-1))}^{\kappa-1}(u)} + \frac{\gamma_{2j+1}}{F_{(k+1,2j)}^{\kappa-1}(u)}}{u_{2j}} + \frac{\frac{\gamma_{2j-1}}{F_{(k+1,2(j-1))}^{\kappa-1}(v)} + \frac{\gamma_{2j+1}}{F_{(k+1,2j)}^{\kappa-1}(v)}}{v_{2j}}}{\frac{\gamma_{2j-1}}{F_{(k+1,2(j-1))}^{\kappa-1}(v)} + \frac{\gamma_{2j+1}}{F_{(k+1,2j)}^{\kappa-1}(v)}} + \frac{\gamma_{2j-1}}{F_{(k+1,2(j-1))}^{\kappa-1}(u)} + \frac{\gamma_{2j+1}}{F_{(k+1,2j)}^{\kappa-1}(u)}}} \right) \end{aligned}$$

By (4.40) and (4.41), we have that $\frac{F_{(k+1,2(j-1))}^{\kappa-1}(v)}{v_{2j}} \leq \frac{F_{(k+1,2(j-1))}^{\kappa-1}(u)}{u_{2j}}$ (recall $\frac{v_{2(j-1)}}{u_{2(j-1)}} =$

$\frac{v_{2j}}{u_{2j}} = \frac{v_2}{u_2}$), $\frac{F_{(k+1,2j)}^{\kappa-1}(v)}{v_{2j}} \leq \frac{F_{(k+1,2j)}^{\kappa-1}(u)}{u_{2j}}$ and

$\frac{F_{(k+1,2(j+1))}^{\kappa-1}(v)}{v_{2j}} \leq \frac{F_{(k+1,2(j+1))}^{\kappa-1}(u)}{u_{2j}}$, with strict inequality for at least one of these terms.

Hence $\frac{F_{(k+1,2j)}^\kappa(v)}{F_{(k+1,2j)}^\kappa(u)} < \frac{v_2}{u_2}$.

□

We can now consider the lemma for the odd case, $m = k + 2$. Observe that for

$$j \in \left\{ 1, \dots, \left\lceil \frac{k+2}{2} \right\rceil - 1 \right\}$$

$$\frac{F_{(k+2,2j)}^{\left\lceil \frac{k+2}{2} \right\rceil - 1}(v_2, \dots, v_{2(k+2)})}{F_{(k+2,2j)}^{\left\lceil \frac{k+2}{2} \right\rceil - 1}(u_2, \dots, u_{2(k+2)})} = \frac{F_{(k+1,2j)}^{\left\lceil \frac{k+1}{2} \right\rceil}(v_2, \dots, v_{2(k+1)})}{F_{(k+1,2j)}^{\left\lceil \frac{k+1}{2} \right\rceil}(u_2, \dots, u_{2(k+1)})} < \frac{v_{2j}}{u_{2j}}$$

The equality follows by the same reasoning as before - the function

$F_{(k+2,2j)}^{\left\lceil \frac{k+2}{2} \right\rceil - 1}(u_2, \dots, u_{2(k+2)})$ does not depend on the variable $u_{2(k+2)}$ and is represented

by the same continued fraction as the function

$F_{(k+1,2j)}^{\left\lceil \frac{k+1}{2} \right\rceil}(u_2, \dots, u_{2(k+1)})$. The inequality follows from (4.42) for the case $m = k + 1$.

Similarly, if once again we let $\tilde{F}_{(k+1,2j)}^{\left\lceil \frac{k+1}{2} \right\rceil}$ be the function $F_{(k+1,2j)}^{\left\lceil \frac{k+1}{2} \right\rceil}$ with

$(\gamma_1^t, \gamma_2^t, \dots, \gamma_{2k+3}^t)$ replaced by $(\gamma_3^t, \gamma_4^t, \dots, \gamma_{2k+5}^t)$, we can conclude that for $j \in \left\{ \left\lceil \frac{(k+2)+1}{2} \right\rceil + 1, \dots, k + 2 \right\}$

$$\frac{F_{(k+2,2j)}^{\left\lceil \frac{k+2}{2} \right\rceil - 1}(v_2, \dots, v_{2(k+2)})}{F_{(k+2,2j)}^{\left\lceil \frac{k+2}{2} \right\rceil - 1}(u_2, \dots, u_{2(k+2)})} = \frac{\tilde{F}_{(k+1,2(j-1))}^{\left\lceil \frac{k+1}{2} \right\rceil}(v_4, \dots, v_{2(k+2)})}{\tilde{F}_{(k+1,2(j-1))}^{\left\lceil \frac{k+1}{2} \right\rceil}(u_4, \dots, u_{2(k+2)})} < \frac{v_2}{u_2}$$

Lastly, if $j = \kappa := \left\lceil \frac{k+2}{2} \right\rceil$, we observe that

$$\frac{F_{(k+2,2\kappa)}^{\kappa-1}(v_2, \dots, v_{2(k+2)})}{F_{(k+2,2\kappa)}^{\kappa-1}(u_2, \dots, u_{2(k+2)})} = \frac{v_{2\kappa}}{u_{2\kappa}} \cdot \left(\frac{\frac{\gamma_{2\kappa-1}}{F_{(k+2,2(\kappa-1))}^{\kappa-2}(u)} + \frac{\gamma_{2\kappa+1}}{F_{(k+2,2(\kappa+1))}^{\kappa-2}(u)}}{\frac{F_{(k+2,2\kappa)}^{\kappa-2}(u)}{u_{2\kappa}} + \frac{F_{(k+2,2\kappa)}^{\kappa-1}(u)}{u_{2\kappa}}} + \frac{\frac{\gamma_{2j-1}}{F_{(k+2,2(\kappa-1))}^{\kappa-2}(v)} + \frac{\gamma_{2j+1}}{F_{(k+2,2(\kappa+1))}^{\kappa-2}(v)}}{\frac{F_{(k+2,2\kappa)}^{\kappa-2}(v)}{v_{2\kappa}} + \frac{F_{(k+2,2\kappa)}^{\kappa-1}(v)}{v_{2\kappa}}} \right)$$

But $\kappa - 2 = \left\lceil \frac{k+1}{2} \right\rceil - 1$, so

$$\frac{F_{(k+2,2(\kappa-1))}^{\kappa-2}(v)}{F_{(k+2,2(\kappa-1))}^{\kappa-2}(u)} = \frac{F_{(k+1,2(\kappa-1))}^{\kappa-2}(v_2, \dots, v_{2(k+1)})}{F_{(k+1,2(\kappa-1))}^{\kappa-2}(u_2, \dots, u_{2(k+1)})} = \frac{v_2}{u_2}$$

and

$$\frac{F_{(k+2,2\kappa)}^{\kappa-2}(v)}{F_{(k+2,2\kappa)}^{\kappa-2}(u)} = \frac{F_{(k+1,2\kappa)}^{\kappa-2}(v_2, \dots, v_{2(k+1)})}{\widehat{F}_{(k+1,2\kappa)}^{\kappa-2}(u_2, \dots, u_{2(k+1)})} = \frac{v_2}{u_2}$$

by the conclusion of equality (4.41) for $m = k + 1$. We conclude by the same reasoning that

$$\frac{F_{(k+2,2(\kappa+1))}^{\kappa-2}(v)}{F_{(k+2,2(\kappa+1))}^{\kappa-2}(u)} = \frac{\widetilde{F}_{(k+1,2\kappa)}^{\kappa-2}(v_4, \dots, v_{2(k+2)})}{\widetilde{F}_{(k+1,2\kappa)}^{\kappa-2}(u_4, \dots, u_{2(k+2)})} = \frac{v_2}{u_2}$$

All this implies that $\left(\frac{\frac{\frac{\gamma_{2j-1}}{F_{2(j-1)}^{\kappa-2}(u)} + \frac{\gamma_{2j-1}}{F_{2j}^{\kappa-2}(u)}}{u_{2j}} + \frac{\frac{\gamma_{2j+1}}{F_{2j}^{\kappa-2}(u)} + \frac{\gamma_{2j+1}}{F_{2(j+1)}^{\kappa-2}(u)}}{u_{2j}}}{\frac{\gamma_{2j-1}}{F_{2(j-1)}^{\kappa-2}(v)} + \frac{\gamma_{2j-1}}{F_{2j}^{\kappa-2}(v)}}{v_{2j}} + \frac{\frac{\gamma_{2j+1}}{F_{2j}^{\kappa-2}(v)} + \frac{\gamma_{2j+1}}{F_{2(j+1)}^{\kappa-2}(v)}}{v_{2j}}} \right) = 1$. This concludes the proof of (4.40) and (4.41) for odd $m = k + 2$, and (4.42) follows by the same reasoning as it did previously for $m = k + 1$. \square

Appendix B

Before proceeding with bounding the two constants C_π and C_J appearing in the proof of Corollary 4.3, we will make the following definition used in deriving an upper bound for (19).

Definition .1. Let $f \in L_1(\mathbb{R})$ be non-negative. Define $\text{med}(f)$ to be the infimum over $m \in \mathbb{R}$ such that

$$\int_{-\infty}^m f = \int_m^{\infty} f$$

Lemma .5. Suppose $f(v) \in L_1(\mathbb{R})$ is non-negative, and $g(v)$ is non-negative, monotone decreasing and $fg(v) \in L_1(\mathbb{R})$. Then $\text{med}(fg) \leq \text{med}(f)$.

Proof. Let $m = \text{med}(f)$. Then

$$\int_{-\infty}^m fg \geq g(m) \int_{-\infty}^m f = g(m) \int_m^{\infty} f \geq \int_m^{\infty} fg$$

Hence $\text{med}(fg) \leq m \leq \text{med}(f)$. □

Corollary .6. Suppose $f(v) \in L_1(\mathbb{R}^+)$ is non-negative, and $\sigma \geq 0$. Then $0 < y_1 \leq y_2$ implies $\text{med}\left(\frac{f(v)}{(v+y_1)^\sigma}\right) \leq \text{med}\left(\frac{f(v)}{(v+y_2)^\sigma}\right)$.

Proof. Since $\left(\frac{v+y_2}{v+y_1}\right)^\sigma$ is monotone decreasing in v and $\frac{f(v)}{(v+y_1)^\sigma} = \frac{f(v)}{(v+y_2)^\sigma} \left(\frac{v+y_2}{v+y_1}\right)^\sigma$, the statement of the corollary is an immediate consequence of Lemma .5. □

Proof of Corollary 4.3. We can bound the ratio

$$\tilde{C}_\pi := \int \left(\frac{\max_i \{1, v_i\}}{\min_i \{1, v_i\}} \right) \left(\prod_{i=1}^4 v_i^{a_i + a_{i+1} - 1} \right) \exp \left(\sum_{i=1}^5 -v_i v_{i-1} \right) dv / C_g \quad (16)$$

appearing in the proof of Corollary 4.3 by first considering the simplification

$$\left(\frac{\max_i \{1, v_i\}}{\min_i \{1, v_i\}} \right) \leq \sum_{i=1}^4 \mathbf{1}_{\{v_i \geq 1\}} v_i + \sum_{i=1}^4 \mathbf{1}_{\{v_i \leq 1\}} \frac{1}{v_i} + \sum_{i \neq j; i, j \leq 4} \mathbf{1}_{\{v_i \geq 1, v_j \leq 1\}} \frac{v_i}{v_j} \quad (17)$$

It would therefore suffice to obtain an upper bound on the sum obtained by substituting (17) in the aforementioned ratio. Throughout this section we assume that the term $N(w_1, w_2, w_3, w_4)$ (to be defined) is finite for values of (w_1, w_2, w_3, w_4) relevant to our computation. This will indeed be confirmed at the end.

Let $N(w_1, w_2, w_3, w_4) := \int \left(\prod_{i=1}^4 v_i^{w_i - 1} \right) \exp \left(\sum_{i=1}^5 -v_i v_{i-1} \right)$, and observe that by integrating w.r.t. v_1 and v_4 we get

$$\begin{aligned} & N(\alpha_1, \alpha_2, \alpha_3, \alpha_4) \\ &= \Gamma(\alpha_1) \Gamma(\alpha_4) \int \frac{v_2^{\alpha_2 - 1}}{(x + v_2)^{\alpha_1}} \frac{v_3^{\alpha_3 - 1}}{(b + v_3)^{\alpha_4}} \exp(-v_2 v_3) dv_2 dv_3 \end{aligned} \quad (18)$$

Hence

$$\begin{aligned}
\frac{N(\alpha_1 + 1, \alpha_2, \alpha_3, \alpha_4)}{N(\alpha_1, \alpha_2, \alpha_3, \alpha_4)} &= \alpha_1 \left(\int \frac{v_2^{\alpha_2-1}}{(x+v_2)^{\alpha_1+1}} \frac{v_3^{\alpha_3-1}}{(b+v_3)^{\alpha_4}} \exp(-v_2 v_3) dv_2 dv_3 \right) \\
&/ \left(\int \frac{v_2^{\alpha_2-1}}{(x+v_2)^{\alpha_1}} \frac{v_3^{\alpha_3-1}}{(b+v_3)^{\alpha_4}} \exp(-v_2 v_3) dv_2 dv_3 \right) \\
&\leq \alpha_1 \left(\int \frac{1}{x} \frac{v_2^{\alpha_2-1}}{(x+v_2)^{\alpha_1}} \frac{v_3^{\alpha_3-1}}{(b+v_3)^{\alpha_4}} \exp(-v_2 v_3) dv_2 dv_3 \right) \\
&/ \left(\int \frac{v_2^{\alpha_2-1}}{(x+v_2)^{\alpha_1}} \frac{v_3^{\alpha_3-1}}{(b+v_3)^{\alpha_4}} \exp(-v_2 v_3) dv_2 dv_3 \right) \\
&= \frac{\alpha_1}{x}
\end{aligned}$$

By symmetry it follows immediately that $N(\alpha_1, \alpha_2, \alpha_3, \alpha_4 + 1) / N(\alpha_1, \alpha_2, \alpha_3, \alpha_4) \leq \frac{\alpha_4}{b}$.

We would now like to consider the ratio $N_3(\alpha_1, \alpha_2, \alpha_3 + 1, \alpha_4) / N(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$, where $N_j(w_1, w_2, w_3, w_4) := \int \left(\mathbf{1}_{\{v_j \geq 1\}} v_j \prod_{i=1}^4 v_i^{w_i-1} \right) \exp\left(\sum_{i=1}^5 -v_i v_{i-1}\right) dv$. Our goal is to arrive at an (good) upper bound for (16). When we substitute (17) in the integral in (16), we obtain (after moving the integral inside the summation) three summations of integrals, corresponding to the three summations in the right-hand side of (17). It follows that for the first summation it would be sufficient to consider the sum of ratios of this form since

$$\begin{aligned}
&\sum_{i=1}^4 \frac{N_i(\alpha_1, \dots, \alpha_i + 1, \dots, \alpha_4)}{N(\alpha_1, \alpha_2, \alpha_3, \alpha_4)} \\
&= \int \left(\sum_{i=1}^4 \mathbf{1}_{\{v_i \geq 1\}} v_i \right) \left(\prod_{i=1}^4 v_i^{\alpha_i + \alpha_{i+1} - 1} \right) \exp\left(\sum_{i=1}^5 -v_i v_{i-1}\right) dv / C_g
\end{aligned}$$

Integrating first in v_2 and v_4 , we get

$$\frac{N_3(\alpha_1, \alpha_2, \alpha_3 + 1, \alpha_4)}{N(\alpha_1, \alpha_2, \alpha_3, \alpha_4)} = \left(\int \frac{v_1^{\alpha_1-1}}{(v_1 + v_3)^{\alpha_2}} \frac{\mathbf{1}_{\{v_3 \geq 1\}} v_3^{\alpha_3}}{(v_3 + b)^{\alpha_4}} \exp(-xv_1) dv_1 dv_3 \right) \quad (19)$$

$$/ \left(\int \frac{v_1^{\alpha_1-1}}{(v_1 + v_3)^{\alpha_2}} \frac{v_3^{\alpha_3-1}}{(v_3 + b)^{\alpha_4}} \exp(-xv_1) dv_1 dv_3 \right)$$

Let

$$g(v_1) := \int \frac{\mathbf{1}_{\{v_3 \geq 1\}} v_3^{\alpha_3}}{(v_1 + v_3)^{\alpha_2} (v_3 + b)^{\alpha_4}} dv_3$$

Recall that $\alpha_i = a_i + a_{i+1}$ and we assume apriori that $a_i \geq 1$, hence $\alpha_2 + \alpha_4 - \alpha_3 = a_2 + a_5$, and g is finite on \mathbb{R}^+ . Furthermore, since g is monotone decreasing, it follows by Lemma .5 and from [9] that

$$\text{med} \left(g(v_1) v_1^{\alpha_1-1} \exp(-xv_1) \right) \leq \text{med} \left(v_1^{\alpha_1-1} \exp(-xv_1) \right) \leq \frac{\alpha_1}{x}$$

Thus

$$\int \frac{v_1^{\alpha_1-1} \exp(-xv_1)}{(v_1 + v_3)^{\alpha_2}} \frac{\mathbf{1}_{\{v_3 \geq 1\}} v_3^{\alpha_3}}{(v_3 + b)^{\alpha_4}} dv_1 dv_3 \leq 2 \int \frac{\mathbf{1}_{\{v_1 \leq \frac{\alpha_1}{x}\}} v_1^{\alpha_1-1} \exp(-xv_1)}{(v_1 + v_3)^{\alpha_2}} \frac{\mathbf{1}_{\{v_3 \geq 1\}} v_3^{\alpha_3}}{(v_3 + b)^{\alpha_4}} dv_1 dv_3 \quad (20)$$

Then by applying Corollary .6 repeatedly, we can conclude (when considering the argument of med as a function of v_3) that

$$\text{med} \left(\frac{\mathbf{1}_{\{v_1 \leq \frac{\alpha_1}{x}\}}}{(v_1 + v_3)^{\alpha_2}} \frac{\mathbf{1}_{\{v_3 \geq 1\}} v_3^{\alpha_3}}{(v_3 + b)^{\alpha_4}} \right) \leq \text{med} \left(\frac{\mathbf{1}_{\{v_3 \geq 1\}} v_3^{\alpha_3}}{\left(\frac{\alpha_1}{x} + v_3\right)^{\alpha_2} (v_3 + b)^{\alpha_4}} \right) \quad (21)$$

$$\leq \text{med} \left(\frac{\mathbf{1}_{\{v_3 \geq 1\}} v_3^{\alpha_3}}{\left(v_3 + \frac{\alpha_1}{x} + b\right)^{\alpha_4 + \alpha_2}} \right)$$

We can bound the right-most term in (21) by the following method: note that

$$\frac{\mathbf{1}_{\{v_3 \geq 1\}} v_3^{\alpha_3}}{\left(v_3 + \frac{\alpha_1}{x} + b\right)^{\alpha_4 + \alpha_2}} = \frac{\mathbf{1}_{\{v_3 \geq 1\}}}{v_3^{\alpha_4 + \alpha_2 - \alpha_3 - 0.5}} \frac{v_3^{\alpha_4 + \alpha_2 - 0.5}}{\left(v_3 + \frac{\alpha_1}{x} + b\right)^{\alpha_4 + \alpha_2}} \text{ and } \frac{v_3^{\alpha_4 + \alpha_2 - 0.5}}{\left(v_3 + \frac{\alpha_1}{x} + b\right)^{\alpha_4 + \alpha_2}} \text{ is decreasing when-}$$

ever

$v \geq 2 \left(\frac{\alpha_1}{x} + b\right) (\alpha_4 + \alpha_2 - 0.5)$. Therefore, by Lemma .5

$$\begin{aligned} & \text{med} \left(\frac{\mathbf{1}_{\{v_3 \geq 1\}} v_3^{\alpha_3}}{\left(v_3 + \frac{\alpha_1}{x} + b\right)^{\alpha_4 + \alpha_2}} \right) \\ & \leq \text{med} \left(\frac{\mathbf{1}_{\left\{v_3 \geq \max\left\{1, 2\left(\frac{\alpha_1}{x} + b\right)(\alpha_4 + \alpha_2 - 0.5)\right\}\right\}} v_3^{\alpha_3}}{\left(v_3 + \frac{\alpha_1}{x} + b\right)^{\alpha_4 + \alpha_2}} \right) \\ & \leq \text{med} \left(\frac{\mathbf{1}_{\left\{v_3 \geq \max\left\{1, 2\left(\frac{\alpha_1}{x} + b\right)(\alpha_4 + \alpha_2 - 0.5)\right\}\right\}}}{v_3^{\alpha_4 + \alpha_2 - \alpha_3 - 0.5}} \right) \\ & = \max \left\{ 1, 2 \left(\frac{\alpha_1}{x} + b\right) (\alpha_4 + \alpha_2 - 0.5) \right\} 2^{\frac{1}{\alpha_2 + \alpha_5 - 1.5}} \end{aligned} \quad (22)$$

Setting $m_1 = \max \left\{ 1, 2 \left(\frac{\alpha_1}{x} + b\right) (\alpha_4 + \alpha_2 - 0.5) \right\} 2^{\frac{1}{\alpha_2 + \alpha_5 - 1.5}}$ and applying (20) and

(21), we conclude

$$\begin{aligned}
& \frac{N_3(\alpha_1, \alpha_2, \alpha_3 + 1, \alpha_4)}{N(\alpha_1, \alpha_2, \alpha_3, \alpha_4)} \\
&= \left(\int \frac{v_1^{\alpha_1-1}}{(v_1 + v_3)^{\alpha_2}} \frac{\mathbf{1}_{\{v_3 \geq 1\}} v_3^{\alpha_3}}{(v_3 + b)^{\alpha_4}} \exp(-xv_1) \right) / \left(\int \frac{v_1^{\alpha_1-1}}{(v_1 + v_3)^{\alpha_2}} \frac{v_3^{\alpha_3-1}}{(v_3 + b)^{\alpha_4}} \exp(-xv_1) \right) \\
&\leq \left(2 \int \frac{\mathbf{1}_{\{v_1 \leq \frac{\alpha_1}{x}\}} v_1^{\alpha_1-1}}{(v_1 + v_3)^{\alpha_2}} \frac{\mathbf{1}_{\{v_3 \geq 1\}} v_3^{\alpha_3}}{(v_3 + b)^{\alpha_4}} \exp(-xv_1) \right) / \\
&\quad \left(\int \frac{v_1^{\alpha_1-1}}{(v_1 + v_3)^{\alpha_2}} \frac{v_3^{\alpha_3-1}}{(v_3 + b)^{\alpha_4}} \exp(-xv_1) \right) \\
&\leq \left(4 \int \frac{\mathbf{1}_{\{v_1 \leq \frac{\alpha_1}{x}\}} v_1^{\alpha_1-1}}{(v_1 + v_3)^{\alpha_2}} \frac{\mathbf{1}_{\{v_3 \geq 1\}} \mathbf{1}_{\{v_3 \leq m_1\}} v_3^{\alpha_3}}{(v_3 + b)^{\alpha_4}} \exp(-xv_1) \right) / \\
&\quad \left(\int \frac{v_1^{\alpha_1-1}}{(v_1 + v_3)^{\alpha_2}} \frac{v_3^{\alpha_3-1}}{(v_3 + b)^{\alpha_4}} \exp(-xv_1) \right) \\
&\leq \left(4m_1 \int \frac{\mathbf{1}_{\{v_1 \leq \frac{\alpha_1}{x}\}} v_1^{\alpha_1-1}}{(v_1 + v_3)^{\alpha_2}} \frac{\mathbf{1}_{\{v_3 \geq 1\}} \mathbf{1}_{\{v_3 \leq m_1\}} v_3^{\alpha_3-1}}{(v_3 + b)^{\alpha_4}} \exp(-xv_1) \right) / \\
&\quad \left(\int \frac{v_1^{\alpha_1-1}}{(v_1 + v_3)^{\alpha_2}} \frac{v_3^{\alpha_3-1}}{(v_3 + b)^{\alpha_4}} \exp(-xv_1) \right) \\
&\leq 4 \max \left\{ 1, 2 \left(\frac{\alpha_1}{x} + b \right) (\alpha_4 + \alpha_2 - 0.5) \right\} 2^{\frac{1}{\alpha_2 + \alpha_5 - 1.5}} \tag{23}
\end{aligned}$$

By the symmetry of $N(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ we can also conclude that

$$\begin{aligned}
& N_2(\alpha_1, \alpha_2 + 1, \alpha_3, \alpha_4) / N(\alpha_1, \alpha_2, \alpha_3, \alpha_4) \\
&\leq 4 \max \left\{ 1, 2 \left(\frac{\alpha_4}{b} + x \right) (\alpha_3 + \alpha_1 - 0.5) \right\} 2^{\frac{1}{\alpha_2 + \alpha_5 - 1.5}}
\end{aligned}$$

Observe next that

$$\begin{aligned}
& \frac{N(\alpha_1 - 1, \alpha_2, \alpha_3, \alpha_4)}{N(\alpha_1, \alpha_2, \alpha_3, \alpha_4)} \\
&= \frac{1}{\alpha_1 - 1} \left(\int \frac{v_2^{\alpha_2-1}}{(x+v_2)^{\alpha_1-1}} \frac{v_4^{\alpha_4-1} \exp(-bv_4)}{(v_2+v_4)^{\alpha_3}} dv_2 dv_4 \right) \\
& \quad / \left(\int \frac{v_2^{\alpha_2-1}}{(x+v_2)^{\alpha_1}} \frac{v_4^{\alpha_4-1} \exp(-bv_4)}{(v_2+v_4)^{\alpha_3+1}} dv_2 dv_4 \right) \\
&= \frac{1}{\alpha_1 - 1} \left(\int \frac{v_2^{\alpha_2-1} (x+v_2)}{(x+v_2)^{\alpha_1}} \frac{v_4^{\alpha_4-1} \exp(-bv_4)}{(v_2+v_4)^{\alpha_3}} dv_2 dv_4 \right) \\
& \quad / \left(\int \frac{v_2^{\alpha_2-1}}{(x+v_2)^{\alpha_1}} \frac{v_4^{\alpha_4-1} \exp(-bv_4)}{(v_2+v_4)^{\alpha_3+1}} dv_2 dv_4 \right) \\
&= \frac{1}{\alpha_1 - 1} \left(x + \frac{N(\alpha_1, \alpha_2 + 1, \alpha_3, \alpha_4)}{N(\alpha_1, \alpha_2, \alpha_3, \alpha_4)} \right) \\
&\leq \frac{1}{\alpha_1 - 1} \left(x + 1 + \frac{N_2(\alpha_1, \alpha_2 + 1, \alpha_3, \alpha_4)}{N(\alpha_1, \alpha_2, \alpha_3, \alpha_4)} \right) \\
&\leq \frac{1}{\alpha_1 - 1} \left(x + 1 + 4 \max \left\{ 1, 2 \left(\frac{\alpha_4}{b} + x \right) (\alpha_3 + \alpha_1 - 0.5) \right\} 2^{\frac{1}{\alpha_2 + \alpha_5 - 1.5}} \right)
\end{aligned}$$

The second-last inequality is a result of the fact that

$$\begin{aligned}
& N(\alpha_1, \alpha_2 + 1, \alpha_3, \alpha_4) \\
&= N_2(\alpha_1, \alpha_2 + 1, \alpha_3, \alpha_4) + \int \left(\mathbf{1}_{\{v_i < 1\}} v_i \prod_{i=1}^4 v_i^{\alpha_i-1} \right) \exp \left(\sum_{i=1}^5 -v_i v_{i-1} \right)
\end{aligned}$$

and the second term in the sum is less than or equal to the denominator $N(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$.

Similarly we obtain

$$\begin{aligned}
& \frac{N(\alpha_1, \alpha_2, \alpha_3, \alpha_4 - 1)}{N(\alpha_1, \alpha_2, \alpha_3, \alpha_4)} \\
&\leq \frac{1}{\alpha_4 - 1} \left(b + 1 + 4 \max \left\{ 1, 2 \left(\frac{\alpha_1}{x} + b \right) (\alpha_4 + \alpha_2 - 0.5) \right\} 2^{\frac{1}{\alpha_2 + \alpha_5 - 1.5}} \right)
\end{aligned}$$

Proceeding in this manner we can also conclude that

$$\frac{N(\alpha_1, \alpha_2 - 1, \alpha_3, \alpha_4)}{N(\alpha_1, \alpha_2, \alpha_3, \alpha_4)} = \frac{1}{\alpha_2 - 1} \left(\frac{N(\alpha_1 + 1, \alpha_2, \alpha_3, \alpha_4)}{N(\alpha_1, \alpha_2, \alpha_3, \alpha_4)} + \frac{N(\alpha_1, \alpha_2, \alpha_3 + 1, \alpha_4)}{N(\alpha_1, \alpha_2, \alpha_3, \alpha_4)} \right) \quad (24)$$

and

$$\frac{N(\alpha_1, \alpha_2, \alpha_3 - 1, \alpha_4)}{N(\alpha_1, \alpha_2, \alpha_3, \alpha_4)} = \frac{1}{\alpha_3 - 1} \left(\frac{N(\alpha_1, \alpha_2, \alpha_3, \alpha_4 + 1)}{N(\alpha_1, \alpha_2, \alpha_3, \alpha_4)} + \frac{N(\alpha_1, \alpha_2 + 1, \alpha_3, \alpha_4)}{N(\alpha_1, \alpha_2, \alpha_3, \alpha_4)} \right)$$

Next we provide an upper bound for the terms pertaining to the sum

$\sum_{i \neq j; i, j \leq 4} \mathbf{1}_{\{v_i \geq 1, v_j \leq 1\}} \frac{v_i}{v_j}$ and its role in the ratio C_π in Corollary 4.3. Note that for

the case $i = 1, j = 2$ this is given by

$$\begin{aligned} & \frac{N(\alpha_1 + 1, \alpha_2 - 1, \alpha_3, \alpha_4)}{N(\alpha_1, \alpha_2, \alpha_3, \alpha_4)} \\ &= \frac{N(\alpha_1 + 1, \alpha_2 - 1, \alpha_3, \alpha_4)}{N(\alpha_1 + 1, \alpha_2, \alpha_3, \alpha_4)} \frac{N(\alpha_1 + 1, \alpha_2, \alpha_3, \alpha_4)}{N(\alpha_1, \alpha_2, \alpha_3, \alpha_4)} \end{aligned} \quad (25)$$

As was shown, the second term in the product on the right-hand side is bounded

from above by $\frac{\alpha_1}{x}$, while the first term is of the same form as the term

$N(\alpha_1, \alpha_2 - 1, \alpha_3, \alpha_4) / N(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ and following an analogous derivation to

(24) we can conclude that it is bounded from above by

$$\begin{aligned} & \left| \frac{N(\alpha_1 + 1, \alpha_2 - 1, \alpha_3, \alpha_4)}{N(\alpha_1, \alpha_2, \alpha_3, \alpha_4)} \right| \\ & \leq \frac{\alpha_1}{x} \frac{1}{\alpha_2 - 1} \left(\frac{N(\alpha_1 + 2, \alpha_2, \alpha_3, \alpha_4)}{N(\alpha_1 + 1, \alpha_2, \alpha_3, \alpha_4)} + \frac{N(\alpha_1 + 1, \alpha_2, \alpha_3 + 1, \alpha_4)}{N(\alpha_1 + 1, \alpha_2, \alpha_3, \alpha_4)} \right) \\ & \leq \frac{\alpha_1}{x} \frac{1}{\alpha_2 - 1} \left(\frac{\alpha_1 + 1}{x} + 4max \left\{ 1, 2 \left(\frac{\alpha_1 + 1}{x} + b \right) (\alpha_4 + \alpha_2 - 0.5) \right\} 2^{\frac{1}{a_2 + a_5 - 1.5}} \right) \end{aligned}$$

A similar derivation follows for other values of i and j .

We can now summarise these results: let $\varphi_1 = \frac{\alpha_1}{x}$,
 $\varphi_2 = 4\max \left\{ 1, 2 \left(\frac{\alpha_4}{b} + x \right) (\alpha_3 + \alpha_1 - 0.5) \right\} 2^{\frac{1}{a_2+a_5-1.5}}$,
 $\varphi_3 = 4\max \left\{ 1, 2 \left(\frac{\alpha_1}{x} + b \right) (\alpha_4 + \alpha_2 - 0.5) \right\} 2^{\frac{1}{a_2+a_5-1.5}}$ and $\varphi_4 = \frac{\alpha_4}{b}$. Also let $\varphi_5 = \frac{1}{\alpha_1-1} (x + 1 + \varphi_2)$, $\varphi_6 = \frac{1}{\alpha_2-1} (\varphi_1 + \varphi_3 + 1)$, $\varphi_7 = \frac{1}{\alpha_3-1} (\varphi_4 + \varphi_2 + 1)$ and $\varphi_8 = \frac{1}{\alpha_4-1} (b + 1 + \varphi_3)$. Lastly let φ_9 be same as φ_5 but with every occurrence of α_1 replaced by $\alpha_1 + 1$, and similar definition follows for φ_{10} , φ_{11} and φ_{12} . Then

$$\int \left(\frac{\max_i \{1, v_i\}}{\min_i \{1, v_i\}} \right) \left(\prod_{i=1}^4 v_i^{a_i+a_{i+1}-1} \right) \exp \left(\sum_{i=1}^5 -v_i v_{i-1} \right) dv / C_g$$

$$\leq \sum_{1 \leq i \leq 8} \varphi_i + \sum_{1 \leq i \leq 4, 9 \leq j \leq 12, j \neq i+8} \varphi_i \varphi_j \quad (26)$$

It remains to verify that $N(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ is in fact finite (the bounds in this

section would then guarantee the finiteness of similar terms).

$$\begin{aligned}
& N(\alpha_1, \alpha_2, \alpha_3, \alpha_4) \\
&= \Gamma(\alpha_2) \Gamma(\alpha_4) \int (\mathbf{1}_{\{v_1+v_3 \leq 1\}} + \mathbf{1}_{\{v_1+v_3 > 1\}}) \frac{v_1^{\alpha_1-1}}{(v_1+v_3)^{\alpha_2}} \frac{v_3^{\alpha_3-1}}{(v_3+b)^{\alpha_4}} \exp(-xv_1) \\
&= \Gamma(\alpha_2) \Gamma(\alpha_4) \left(\int \frac{\mathbf{1}_{\{v_1+v_3 \leq 1\}} v_1^{\alpha_1-1} v_3^{\alpha_3-1} \exp(-xv_1)}{(v_1+v_3)^{\alpha_2} (v_3+b)^{\alpha_4}} + \right. \\
&\quad \left. \int \left(\int \frac{\mathbf{1}_{\{v_1+v_3 > 1\}} v_3^{\alpha_3-1}}{(v_1+v_3)^{\alpha_2} (v_3+b)^{\alpha_4}} dv_3 \right) v_1^{\alpha_1-1} \exp(-xv_1) dv_1 \right) \\
&\leq \Gamma(\alpha_2) \Gamma(\alpha_4) \left(\int \frac{\mathbf{1}_{\{v_1+v_3 \leq 1\}} v_1^{\alpha_1-1} v_3^{\alpha_3-1} \exp(-xv_1)}{(v_1+v_3)^{\alpha_2} (v_3+b)^{\alpha_4}} + \right. \\
&\quad \left. \frac{\Gamma(\alpha_1)}{x^{\alpha_1}} \left(\int_0^1 \frac{v_3^{\alpha_3-1} dv_3}{(v_3+b)^{\alpha_4}} + \int_1^\infty \frac{dv_3}{v_3^{a_2+a_5+1}} \right) \right) \\
&\leq \Gamma(\alpha_2) \Gamma(\alpha_4) \left(\int \frac{\mathbf{1}_{\{v_1+v_3 \leq 1\}}}{(v_3+b)^{\alpha_4}} \exp(-xv_1) + \frac{\Gamma(\alpha_1)}{x^{\alpha_1}} \left(\frac{1}{b^{\alpha_4}} + \int_1^\infty \frac{1}{v_3^{a_2+a_5+1}} \right) \right) < \infty
\end{aligned}$$

The last inequality follows from the fact that on $\{v_1 + v_3 \leq 1\}$ we have $(v_1 + v_3)^{\alpha_2} \geq (v_1 + v_3)^{\alpha_1 + \alpha_3 - 2} \geq v_1^{\alpha_1 - 1} v_3^{\alpha_3 - 1}$.

The final bound is for C_J , which follows easily from the previous derivations.

Recall that $\mathcal{U}^0 = (1, 1, 1, 1)$ and $\mathcal{V}^0 \sim \pi$, while $u_i^0 = \min\{\mathcal{U}_i^0, \mathcal{V}_i^0\}$ and $v_i^0 =$

$\max \{\mathcal{U}_i^0, \mathcal{V}_i^0\}$, and note that

$$\begin{aligned}
\mathbb{E}[J_0] &= \mathbb{E}[K_{1,0} + K_{2,0}] \\
&= \mathbb{E}\left[u_2^0 + u_4^0 + \frac{u_3^0 + u_1^0 + b}{u_2^0(u_3^0 + u_1^0) + u_4^0(u_3^0 + b)}\right] \leq \mathbb{E}\left[u_2^0 + u_4^0 + \frac{1}{u_2^0} + \frac{1}{u_4^0}\right] \\
&\leq \mathbb{E}\left[2 + \mathcal{V}_2^0 + \mathcal{V}_4^0 + 2 + \frac{1}{\mathcal{V}_2^0} + \frac{1}{\mathcal{V}_4^0}\right] \\
&\leq 4 + \varphi_2 + \varphi_4 + \varphi_6 + \varphi_8 \tag{27}
\end{aligned}$$

Setting C_J equal to the right-hand side of the final inequality in (27) completes the proof of Corollary 4.3. □