ANNALES
DE LA FACULTÉ
DES SCIENCES
DE
TOULOUSE
Mathématiques

JAMES P. HOBERT AND KSHITIJ KHARE

*Computable upper bounds on the distance to stationarity for Jovanovski and Madras's Gibbs sampler*

Tome XXIV, n° 4 (2015), p. 935-947.

<http://afst.cedram.org/item?id=AFST_2015_6_24_4_935_0>

Annales de la Faculté des Sciences de Toulouse Vol. XXIV, n° 4, 2015
pp. 935-947

# Computable upper bounds on the distance to stationarity for Jovanovski and Madras's Gibbs sampler

James P. Hobert[(1)], Kshitij Khare[(1)]

**Résumé.** — Une borne supérieure est obtenue pour la distance de Wasserstein à la stationnarité pour une classe de chaînes de Markov sur $\mathbb{R}$. Ce résultat, qui est une généralisation du théorème 2.2 de Diaconis et al. (2009), est appliqué à l'échantillonneur de Gibbs introduit et analysé par Jovanovski et Madras (2014). La borne de Wasserstein qui en résulte est transformée en une borne en variation totale (en utilisant des résultats de Madras et Sezer (2010)), qui est ensuite comparée à une autre borne obtenue par Jovanovski et Madras (2014).

**Abstract.** — An upper bound on the Wasserstein distance to stationarity is developed for a class of Markov chains on $\mathbb{R}$. This result, which is a generalization of Diaconis et al.'s (2009) Theorem 2.2, is applied to a Gibbs sampler Markov chain that was introduced and analyzed by Jovanovski and Madras (2014). The resulting Wasserstein bound is converted into a total variation bound (using results from Madras and Sezer (2010)), and the total variation bound is compared to an alternative bound derived by Jovanovski and Madras (2014).

## 1. Introduction

We begin by describing a Markov chain that was introduced in [5]. Fix positive constants $\{a_i\}_{i=1}^4$, $z_0$ and $z_4$, and consider the probability density function (with respect to Lebesgue measure on $\mathbb{R}^3$) given by

$$\pi(w,x,y) = \qquad (1.1)$$
$$c w^{a_1+a_2-1} x^{a_2+a_3-1} y^{a_3+a_4-1} e^{-z_0 w - wx - xy - y z_4} I_{\mathbb{R}_+}(w) I_{\mathbb{R}_+}(x) I_{\mathbb{R}_+}(y),$$

(1) Department of Statistics, University of Florida

where $c$ is the normalizing constant, and $\mathbb{R}_+ := (0, \infty)$. The density (1.1) arises as the posterior density of a four-level Bayesian hierarchical model that we now describe. Suppose that $(Z_0|W = w, X = x, Y = y) \sim \text{Gamma}(a_1, w)$, $(W|X = x, Y = y) \sim \text{Gamma}(a_2, x)$, $(X|Y = y) \sim \text{Gamma}(a_3, y)$, and, finally, $Y \sim \text{Gamma}(a_4, z_4)$. It's easy to see that (1.1) is the posterior density of the parameter $(W, X, Y)$, given the data $Z_0 = z_0$.

Assume that $(W, X, Y) \sim \pi(w, x, y)$, and note that $W$ and $Y$ are conditionally independent given $X$. In fact, $(W \mid X = x) \sim \text{Gamma}(a_1+a_2, z_0+x)$ and $(Y \mid X = x) \sim \text{Gamma}(a_3 + a_4, x + z_4)$, and $(X \mid W = w, Y = y) \sim \text{Gamma}(a_2 + a_3, w + y)$. Let $g_{\alpha,\beta}(t)$ denote the $\text{Gamma}(\alpha, \beta)$ density evaluated at $t$, i.e.,

$$g_{\alpha,\beta}(t) = \frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\beta t} I_{\mathbb{R}_+}(t) \ .$$

Consider a block Gibbs sampler Markov chain $\{(W_n, Y_n), X_n\}_{n=0}^\infty$ with state space $\mathbb{R}_+^3$ and Markov transition density (Mtd) given by

$$\begin{aligned} k\big((w', y'), x' \mid (w, y), x\big) &= \pi(x'|w', y')\, \pi(w', y'|x) \\ &= g_{a_2+a_3, w'+y'}(x') g_{a_3+a_4, x+z_4}(y') g_{a_1+a_2, x+z_0}(w') \ . \end{aligned} \tag{1.2}$$

It is a *block* Gibbs sampler because $W$ and $Y$ form a block of variables that are updated jointly. Routine arguments show that this chain is Harris ergodic with $\pi(w, x, y)$ as its (unique) invariant density. Note that the right-hand side of (1.2) does not depend on $(w, y)$. It follows that the marginal sequence, $\{X_n\}_{n=0}^\infty$, is itself a Markov chain, and its Mtd is

$$k_G(x' \mid x) = \int_{\mathbb{R}_+^2} g_{a_2+a_3, w+y}(x') g_{a_3+a_4, x+z_4}(y) g_{a_1+a_2, x+z_0}(w)\, dw\, dy \ .$$

This marginal chain is equivalent to the chain defined at the beginning of Section 7 in [5]. It is Harris ergodic and its invariant density is $\pi_G(x) := \int_{\mathbb{R}_+^2} \pi(w, x, y)\, dw\, dy$. The chains driven by $k$ and $k_G$ converge to stationarity (in total variation distance) at exactly the same rate (see, for example, [2] and [8]).

Jovanovski and Madras (2014) (hereafter J&M) used a highly technical one-shot coupling argument to develop an upper bound on the total variation distance to stationarity for the chain $\{X_n\}_{n=0}^\infty$. While it is difficult to give a concise statement of their bound due to its complexity, J&M do state their bound for a particular special case, which we now describe. Fix $z_0 = 1$, $z_4 = 2$, and $a_i = i$ for $i = 1, 2, 3, 4$, and consider starting the chain at the point $X_0 = x_0 = 1$. In this particular situation, their bound simplifies to the following

$$\|P_G^n(1,\cdot) - \Pi_G(\cdot)\|_{\text{TV}} \leqslant 600 \left(\frac{78}{79}\right)^{\frac{n-2}{20}} + 6 \left(\frac{7}{9}\right)^{\lfloor \frac{n}{2} \rfloor + 2}, \tag{1.3}$$

where $P_G^n(x,\cdot)$ is the $n$-step Markov transition function (Mtf) for the chain started at $x \in \mathbb{R}_+$, and $\Pi_G(\cdot)$ is the invariant measure, i.e., $\Pi_G(A) = \int_A \pi_G(x)\,dx$. (We note that there is a typo in J&M: $1 - \frac{78}{79}$ should be $\frac{78}{79}$.) Equation (1.3) implies that roughly 28,500 steps are sufficient for the total variation distance to be less than $10^{-5}$.

In this paper, we perform a new analysis of J&M's Gibbs sampler using a much simpler coupling argument that leads to an upper bound on the *Wasserstein* distance to stationarity. We then use a result from [6] to *convert* this Wasserstein bound into a total variation bound. The resulting total variation bound, which is given in Section 2, has a very simple form, and can give results that are orders of magnitude better than those of J&M. Indeed, for the specific example described above, our bound becomes

$$\|P_G^n(1,\cdot) - \Pi_G(\cdot)\|_{\text{TV}} \leqslant \left\lceil \frac{273}{10} \right\rceil \left(\frac{5}{6}\right)^n,$$

which implies that only 82 steps are enough to get the total variation distance below $10^{-5}$.

To be fair, the chain we consider here is actually the simplest (and only univariate) member of the family of chains analyzed by J&M. While J&M's one-shot coupling technique works for all members of their family, it's not clear whether our technique will extend to handle the other chains. On the other hand, each member of their family has an associated set of parameters (analogous to $\{a_i\}_{i=1}^4$, $z_0$, $z_4$), and the conditions that J&M require on the parameters (in order to get a total variation bound) become more complex and restrictive as the dimension grows.

Finally, we note that the problem we study here is not practically relevant in the sense that the Gibbs sampler described above would likely not be used in practice to explore the intractable density $\pi(w,x,y)$. Indeed, note that

$$\pi_G(x) = \frac{c' x^{a_2+a_3-1}}{(z_0+x)^{a_1+a_2}(z_4+x)^{a_3+a_4}} I_{\mathbb{R}_+}(x),$$

where $c'$ is another normalizing constant. It would not be difficult to design a rejection sampler to make iid draws from this univariate density. Of course, given an exact draw from $\pi_G(x)$, we can get an exact draw from $\pi(w,x,y)$ by simulating two additional independent gamma variates. Thus, classical

Monte Carlo methods could be employed to study $\pi(w,x,y)$, and hence more complex Markov chain Monte Carlo methods, such as the Gibbs sampler, would not be required. Despite this, a convergence rate analysis of the Gibbs sampler is still an interesting endeavor!

The remainder of this paper is organized as follows. In Section 2, we generalize Theorem 2.2 in [3]. The original result provides an upper bound on the Wasserstein distance to stationarity for stochastically monotone Markov chains for which an exact eigen-solution is available. Our version of the theorem does not require an exact eigen-solution. The heart of the paper is Section 3. There we apply the new result to the chain of J&M (which has no obvious exact eigen-solutions), and we convert the resulting Wasserstein bound into a total variation bound. Finally, in Section 4, we briefly compare our Wasserstein bound to an alternative bound derived by [9].

## 2. A bound on the Wasserstein distance to stationarity

Fix $\mathsf{X} \subset \mathbb{R}$ and let $\mathcal{B}(\mathsf{X})$ denote the Borel sets in $\mathsf{X}$. Let $P : \mathsf{X} \times \mathcal{B}(\mathsf{X}) \to [0,1]$ be a Mtf, and assume that the Markov chain defined by $P$ is ergodic with invariant (probability) measure $\Pi(\cdot)$. (For background and definitions, see [7].) Let $P^n$ denote the $n$-step Mtf, where, as usual, $P^1 \equiv P$. We are interested in bounding the distance between the probability measures $P^n(x,\cdot)$ and $\Pi(\cdot)$.

Suppose that $\mu$ and $\nu$ are probability measures on the the real line. The $L^1$-Wasserstein distance, $d_W$, between $\mu$ and $\nu$ is defined as

$$d_W(\mu,\nu) = \inf_{Y\sim\mu, Z\sim\nu} E|Y-Z| \,,$$

where the infimum is taken over all joint distributions for $(Y,Z)$ such that the marginals of $Y$ and $Z$ are $\mu$ and $\nu$, respectively. Here is our first result.

THEOREM 2.1. — *Consider the ergodic Markov chain on $\mathsf{X}$ described above whose invariant measure is $\Pi$. For each fixed $(u_0,v_0) \in \mathsf{X}\times\mathsf{X}$ with $u_0 < v_0$, suppose there exist coupled copies of the Markov chain, $\{U_n\}_{n=0}^\infty$ and $\{V_n\}_{n=0}^\infty$, such that $(U_0,V_0) = (u_0,v_0)$ and $U_n \leqslant V_n$ for all $n \in \{1,2,3,\dots\}$. Suppose further that there exist an increasing function $g : \mathsf{X} \to \mathbb{R}$ and $\lambda \in [0,1)$ (neither depending on $(u_0,v_0)$) such that, for all $n \in \mathbb{N}$ and all $u < v$,*

$$E\big[g(V_n) - g(U_n) \,\big|\, V_{n-1} = v, U_{n-1} = u\big] \leqslant \lambda\big[g(v) - g(u)\big] \,. \tag{2.1}$$

*Assume that*

$$c := \inf_{(u,v)\in\mathsf{X}\times\mathsf{X}, u<v} \frac{g(v)-g(u)}{v-u} > 0 \,.$$

*Then for* $x \in \mathsf{X}$ *and* $n \in \mathbb{N}$*,*

$$d_W\big(P^n(x,\cdot),\Pi(\cdot)\big) \leqslant \frac{\lambda^n}{c} E\big|g(Z)-g(x)\big|\,,$$

*where* $Z \sim \Pi$*.*

*Remark* 2.2.— Theorem 2.1 can be regarded as a generalization of Theorem 2.2 in [3]. In particular, while the conclusions of the two theorems are exactly the same, it's not hard to see that the hypotheses of Theorem 2.1 are weaker than those of Theorem 2.2 in [3]. Indeed, if the Markov chain is stochastically monotone, then the proof of Theorem 2.2 in [3] shows that the coupled chains required by Theorem 2.1 do indeed exist. It then follows that if equation (2.2) in [3] is satisfied, then so is our equation 2.1 (with the same $\lambda$).

*Proof of Theorem 2.1.*— Using the definitions of the Wasserstein distance and the constant $c$, we have

$$\begin{aligned} d_W\big(P^n(u_0,\cdot),P^n(v_0,\cdot)\big) &\leqslant E\big[V_n-U_n \,\big|\, V_0=v_0, U_0=u_0\big] \\ &\leqslant \frac{1}{c} E\big[g(V_n)-g(U_n)\,\big|\, V_0=v_0,U_0=u_0\big]\,. \end{aligned}$$

Now, iterated expectation and application of (2.1) leads to

$$\begin{aligned} &E\big[g(V_n)-g(U_n)\big|V_0=v,U_0=u\big] \\ &\qquad = E\Big[E\big[g(V_n)-g(U_n)\,\big|\,V_{n-1},U_{n-1}\big]\,\Big|\,V_0=v_0,U_0=u_0\Big] \\ &\qquad \leqslant \lambda E\big[g(V_{n-1})-g(U_{n-1})\big|V_0=v_0,U_0=u_0\big]\,, \end{aligned}$$

and repeated application of this argument yields

$$d_W\big(P^n(u_0,\cdot),P^n(v_0,\cdot)\big) \leqslant \frac{\lambda^n}{c}\big[g(v_0)-g(u_0)\big]\,. \tag{2.2}$$

If $u_0 \geqslant v_0$, by exchanging the roles of $u$ and $v$ in the above argument, we get that

$$d_W\big(P^n(u_0,\cdot),P^n(v_0,\cdot)\big) \leqslant \frac{\lambda^n}{c}\big[g(u_0)-g(v_0)\big]\,. \tag{2.3}$$

Combining (2.2) and (2.3), we have that for arbitrary $(u_0,v_0) \in \mathsf{X}\times\mathsf{X}$,

$$d_W\big(P^n(u_0,\cdot),P^n(v_0,\cdot)\big) \leqslant \frac{\lambda^n}{c}\big|g(v_0)-g(u_0)\big|\,. \tag{2.4}$$

Now let $\mathcal{L}$ denote the set of functions $\phi : \mathsf{X} \to \mathbb{R}$ such that $|\phi(x)-\phi(y)| \leqslant |x-y|$ for all $x,y \in \mathsf{X}$, and consider the alternative characterization of $d_W$ given by

$$d_W(\mu,\nu) = \sup_{\phi\in\mathcal{L}} \Big|\int \phi\, d\mu - \int \phi\, d\nu\Big| \tag{2.5}$$

(see, e.g., [4]). If $Z \sim \Pi$, then it follows from (2.5) that

$$d_W\big(P^n(u_0,\cdot),\Pi(\cdot)\big) = d_W\Big(P^n(u_0,\cdot), E\big[P^n(Z,\cdot)\big]\Big) \leqslant E\Big[d_W\big(P^n(u_0,\cdot),P^n(Z,\cdot)\big)\Big]\,.$$

Finally, it follows from (2.4) that

$$d_W\big(P^n(u_0,\cdot),\Pi(\cdot)\big) \leqslant \frac{\lambda^n}{c} E\big|g(Z)-g(u_0)\big|\,.$$

□

## 3. Analysis of Jovanovski and Madras's chain

We begin by constructing coupled copies of the Markov chain driven by $k_G$. Assume that $z_0 \leqslant z_4$. (The case $z_0 > z_4$ can be easily handled using similar arguments.) Fix $(u_0, v_0) \in \mathbb{R}_+^2$ with $u_0 < v_0$. Let $\{G_{1,n}\}_{n=1}^\infty$, $\{G_{2,n}\}_{n=1}^\infty$, and $\{G_{3,n}\}_{n=1}^\infty$ be mutually independent iid sequences of Gamma$(a_1+a_2,1)$, Gamma$(a_3+a_4,1)$, and Gamma$(a_2+a_3,1)$ random variables, respectively. Now, define $\{U_n\}_{n=0}^\infty$ and $\{V_n\}_{n=0}^\infty$ as follows. Set $(U_0,V_0)=(u_0,v_0)$, and, for all $n \in \mathbb{N}$, define

$$U_n = \frac{G_{3,n}}{\frac{G_{1,n}}{z_0+U_{n-1}} + \frac{G_{2,n}}{z_4+U_{n-1}}} \quad \text{and} \quad V_n = \frac{G_{3,n}}{\frac{G_{1,n}}{z_0+V_{n-1}} + \frac{G_{2,n}}{z_4+V_{n-1}}}\,.$$

It's easy to see that $U_n \leqslant V_n$ for all $n$. J&M used these same coupled chains in their analysis.

In order to apply Theorem 2.1, we must establish (2.1). Let $G_1$, $G_2$ and $G_3$ be independent random variables with obvious distributions. Also, let $c = (z_4+u)/(z_0+u)$ and $d=(z_4+v)/(z_0+v)$. Then we have

$$\begin{aligned}
&E\big[V_n - U_n \,\big|\, V_{n-1}=v, U_{n-1}=u\big] \\
&= E\left[\frac{G_3}{\frac{G_1}{z_0+v}+\frac{G_2}{z_4+v}} - \frac{G_3}{\frac{G_1}{z_0+u}+\frac{G_2}{z_4+u}}\right] \\
&= E[G_3]E\left[\frac{(z_0+v)(z_4+v)}{G_1(z_4+v)+G_2(z_0+v)} - \frac{(z_0+u)(z_4+u)}{G_1(z_4+u)+G_2(z_0+u)}\right] \\
&= (a_2+a_3)(v-u)E\left[\frac{(z_4+u)(z_4+v)G_1+(z_0+u)(z_0+v)G_2}{\big(G_1(z_4+u)+G_2(z_0+u)\big)\big(G_1(z_4+v)+G_2(z_0+v)\big)}\right] \\
&= (a_2+a_3)(v-u)E\left[\frac{cdG_1+G_2}{\big(cG_1+G_2\big)\big(dG_1+G_2\big)}\right]\,. \qquad (3.1)
\end{aligned}$$

Note that $c, d \in [1, k]$, where $k := z_4/z_0$. Now,

$$\frac{cdG_1 + G_2}{\big(cG_1 + G_2\big)\big(dG_1 + G_2\big)} = \frac{1}{dG_1 + G_2}\left[d - \frac{(d-1)G_2}{cG_1 + G_2}\right] ,$$

and hence, this quantity is an increasing function of $c$. By symmetry, it's also increasing in $d$. Consequently,

$$\frac{cdG_1 + G_2}{\big(cG_1 + G_2\big)\big(dG_1 + G_2\big)} \leqslant \frac{k^2 G_1 + G_2}{\big(kG_1 + G_2\big)^2} .$$

Therefore,

$$E\big[V_n - U_n \,\big|\, V_{n-1} = v, U_{n-1} = u\big] \leqslant \lambda_G (v - u) , \tag{3.2}$$

where

$$\lambda_G = (a_2 + a_3)\, E\left[\frac{k^2 G_1 + G_2}{\big(kG_1 + G_2\big)^2}\right] .$$

Of course, Theorem 2.1 is not applicable unless $\lambda_G < 1$, and, unfortunately, this is not always the case. In fact, $\lambda_G$ is not always finite. Indeed,

$$\frac{k^2 G_1 + G_2}{\big(kG_1 + G_2\big)^2} \geqslant \frac{kG_1 + G_2}{\big(kG_1 + G_2\big)^2} = \frac{1}{kG_1 + G_2} \geqslant \frac{1}{k\big(G_1 + G_2\big)} ,$$

and $E\big[1/\big(G_1 + G_2\big)\big] = \infty$ if $a_1 + a_2 + a_3 + a_4 \leqslant 1$. (We note that J&M's result also fails in this case.) The following result provides some useful upper bounds on $\lambda_G$.

Proposition 3.1. — *If $a_1 > a_3 + 1$, then*

$$\lambda_G \leqslant \frac{a_2 + a_3}{a_1 + a_2 - 1} < 1 .$$

*If $k \leqslant 2$ and $a_4 > a_2 + 1$, then*

$$\lambda_G \leqslant \frac{a_2 + a_3}{a_3 + a_4 - 1} < 1 .$$

*Proof.* — We provide a proof for the first part. The second part is is proved in a similar way. Fix $g_1 \in (0, \infty)$, $k \geqslant 1$, and define $h : [0, \infty) \to (0, \infty)$ by

$$h(g_2) = \frac{k^2 g_1 + g_2}{\big(kg_1 + g_2\big)^2} .$$

It's easy to see that $h$ is decreasing, so $h(g_2) \leqslant h(0) = 1/g_1$. Thus, assuming $a_1 + a_2 > 1$,

$$\lambda_G = (a_2 + a_3) E\left[ \frac{k^2 G_1 + G_2}{\big(kG_1 + G_2\big)^2} \right] \leqslant (a_2 + a_3) E\big[G_1^{-1}\big] = \frac{a_2 + a_3}{a_1 + a_2 - 1} \ ,$$

and the result follows immediately. □

Assuming that $\{a_i\}_{i=1}^4$ and $k$ are such that $\lambda_G < 1$, Theorem 2.1 implies that, for $x \in \mathbb{R}_+$ and $n \in \mathbb{N}$, we have

$$d_W\big(P^n(x, \cdot), \Pi(\cdot)\big) \leqslant \lambda_G^n \, E\big|Z - x\big| \ ,$$

where $Z \sim \pi_G$. We now establish an upper bound for $E\big|Z - x\big|$ using a standard Lyapunov (or drift) function argument. First, since $Z$ and $x$ are both positive, we have $E\big|Z - x\big| \leqslant E[Z] + x$. Now fix $u \in \mathbb{R}_+$, and assume that $a_1 + a_4 > 1$. Then we have

$$\begin{aligned}
&\mathrm{E}\big[X_{n+1} \,|\, X_n = u\big] \\
&= \int_{\mathbb{R}_+} x' \left[ \int_{\mathbb{R}_+^2} g_{a_2+a_3, w+y}(x')\, g_{a_3+a_4, u+z_4}(y)\, g_{a_1+a_2, u+z_0}(w)\, dw\, dy \right] dx' \\
&= (a_2 + a_3) \int_{\mathbb{R}_+^2} \left( \frac{1}{w + y} \right) g_{a_3+a_4, u+z_4}(y)\, g_{a_1+a_2, u+z_0}(w)\, dw\, dy \\
&\leqslant (a_2 + a_3)\big(u + \max\{z_0, z_4\}\big) \int_{\mathbb{R}_+^2} \frac{g_{a_3+a_4, u+z_4}(y) g_{a_1+a_2, u+z_0}(w)}{(z_0 + u)w + (z_4 + u)y}\, dw\, dy \\
&= \frac{(a_2 + a_3)(u + z_4)}{a_1 + a_2 + a_3 + a_4 - 1} \\
&= \gamma u + L \ , \qquad (3.3)
\end{aligned}$$

where

$$\gamma = \frac{a_2 + a_3}{a_1 + a_2 + a_3 + a_4 - 1} \quad \text{and} \quad L = \frac{z_4(a_2 + a_3)}{a_1 + a_2 + a_3 + a_4 - 1} \ .$$

Note that $\gamma < 1$ since $a_1 + a_4 > 1$. Now, integrating both sides of (3.3) with respect to the invariant density, $\pi_G(u)$, yields $E[Z] \leqslant \gamma E[Z] + L$, or, equivalently $E[Z] \leqslant \frac{L}{1-\gamma}$. (Note that $E[Z] < \infty$ since $a_1 + a_4 > 1$.) Putting all of this together, we have

$$E\big|Z - x\big| \leqslant E[Z] + x \leqslant \frac{z_4(a_2 + a_3)}{a_1 + a_4 - 1} + x \ .$$

*Remark* 3.2. — Note that if $z_0 = z_4$ and $a_1 + a_4 > 1$, then the arguments above show that

$$\mathrm{E}\big[X_{n+1}\,|\,X_n = u\big] = \gamma u + L\ ,$$

from which we see that $E[Z] = \frac{L}{1-\gamma}$. It follows that the function $e(x) = x - E[Z]$ is an eigen-function with eigenvalue $\gamma$; i.e.,

$$\mathrm{E}\big[e(X_{n+1})\,|\,X_n = u\big] = \gamma\, e(u)\ .$$

However, if $z_0 \neq z_4$, there is no obvious eigen-solution.

We now formally state our Wasserstein bound for J&M's chain.

PROPOSITION 3.3. — *Assume that $z_0 \leqslant z_4$, $a_1 + a_4 > 1$, and $\lambda_G < 1$. Then for $x \in \mathbb{R}_+$ and $n \in \mathbb{N}$, we have*

$$d_W\big(P_G^n(x,\cdot), \Pi_G(\cdot)\big) \leqslant \lambda_G^n\, E\big|Z - x\big| \leqslant \lambda_G^n \left[\frac{z_4(a_2 + a_3)}{a_1 + a_4 - 1} + x\right]. \tag{3.4}$$

In order to compare our results with the total variation bounds in J&M, we use a result from [6] to convert our Wasserstein bound into a total variation bound. Here is the conversion result.

PROPOSITION 3.4. — *If $a_1 + a_2 \geqslant 1$ and $a_3 + a_4 \geqslant 1$, then, for all $n \in \mathbb{N}$,*

$$\|P_G^n(x,\cdot) - \Pi_G(\cdot)\|_{TV} \leqslant A\, d_W\big(P_G^{n-1}(x,\cdot), \Pi_G(\cdot)\big)\ ,$$

*where*

$$A = \left(\frac{a_1 + a_2}{z_0} + \frac{a_3 + a_4}{z_4}\right).$$

*Proof.* — According to Theorem 12 in [6], it is enough to show that, for all $u, v \in \mathbb{R}_+$,

$$\int_0^\infty \big|k_G(x|u) - k_G(x|v)\big|\, dx \leqslant 2A|u - v|\ .$$

Now

$$\begin{aligned}
&\big|k_G(x|u) - k_G(x|v)\big| \\
&= \left|\int_{\mathbb{R}_+^2} g_{a_2+a_3,w+y}(x)\Big[g_{a_3+a_4,u+z_4}(y)g_{a_1+a_2,u+z_0}(w) - g_{a_3+a_4,v+z_4}(y)g_{a_1+a_2,v+z_0}(w)\Big]\, dw\, dy\right| \\
&\leqslant \int_{\mathbb{R}_+^2} g_{a_2+a_3,w+y}(x)\Big|g_{a_3+a_4,u+z_4}(y)g_{a_1+a_2,u+z_0}(w) - g_{a_3+a_4,v+z_4}(y)g_{a_1+a_2,v+z_0}(w)\Big|\, dw\, dy\ .
\end{aligned}$$

It follows that

$$\begin{aligned}&\int_{\mathbb{R}_+} \big|k_G(x|u) - k_G(x|v)\big|\,dx \\ &\leqslant \int_{\mathbb{R}_+^2} \Big|g_{a_3+a_4,u+z_4}(y)g_{a_1+a_2,u+z_0}(w) - g_{a_3+a_4,v+z_4}(y)g_{a_1+a_2,v+z_0}(w)\Big|\,dw\,dy\;.\end{aligned}$$

Adding and subtracting $g_{a_3+a_4,u+z_4}(y)g_{a_1+a_2,v+z_0}(w)$ inside the absolute value, and applying the triangle inequality yields

$$\begin{aligned}\int_{\mathbb{R}_+} &\big|k_G(x|u) - k_G(x|v)\big|\,dx \\ &\leqslant \int_{\mathbb{R}_+} \Big|g_{a_1+a_2,u+z_0}(w) - g_{a_1+a_2,v+z_0}(w)\Big|\,dw \\ &\qquad + \int_{\mathbb{R}_+} \Big|g_{a_3+a_4,u+z_4}(y) - g_{a_3+a_4,v+z_4}(y)\Big|\,dy\;. \qquad (3.5)\end{aligned}$$

Now, let $\alpha > 0$ and let $\beta_1 > \beta_2 > 0$. A simple calculation shows that $g_{\alpha,\beta_1}(t) > g_{\alpha,\beta_2}(t)$ if and only if

$$t < t^* := \frac{\alpha}{\beta_1 - \beta_2} \log \frac{\beta_1}{\beta_2}\;.$$

Now fix $\alpha \geqslant 1$ and let $Z \sim \text{Gamma}(\alpha, 1)$. Then

$$\begin{aligned}\int_{\mathbb{R}_+} \big|g_{\alpha,\beta_1}(t) - g_{\alpha,\beta_2}(t)\big|\,dt &= 2\bigg[\int_0^{t^*} g_{\alpha,\beta_1}(t)\,dt - \int_0^{t^*} g_{\alpha,\beta_2}(t)\,dt\bigg] \\ &= 2\Pr\big(\beta_2 t^* < Z < \beta_1 t^*\big) \\ &\leqslant 2t^*(\beta_1 - \beta_2)\bigg[\sup_{t\in\mathbb{R}_+} \frac{e^{-t}t^{\alpha-1}}{\Gamma(\alpha)}\bigg] \\ &= 2t^*(\beta_1 - \beta_2)\bigg[\frac{e^{-(\alpha-1)}(\alpha-1)^{\alpha-1}}{\Gamma(\alpha)}\bigg] \\ &\leqslant 2\alpha \log \frac{\beta_1}{\beta_2}\;, \qquad (3.6)\end{aligned}$$

where the last inequality is due to the fact that, for $s > 0$, $\Gamma(s+1) \geqslant e^{-s}s^s$ (see, e.g., Batir, 2008). Applying (3.6) and then the mean value theorem, we have that

$$\begin{aligned}\int_{\mathbb{R}_+} \Big| g_{a_1+a_2,u+z_0}(w) - g_{a_1+a_2,v+z_0}(w) \Big| \, dw \\ &\leqslant 2(a_1+a_2) \log \left[ \frac{z_0 + \max\{u,v\}}{z_0 + \min\{u,v\}} \right] \\ &= 2(a_1+a_2) \big[ \log \big( z_0 + \max\{u,v\} \big) - \log \big( z_0 + \min\{u,v\} \big) \big] \\ &\leqslant 2(a_1+a_2) \frac{|v-u|}{z_0} \, . \end{aligned} \tag{3.7}$$

Similarly,

$$\int_{\mathbb{R}_+} \Big| g_{a_3+a_4,u+z_4}(y) - g_{a_3+a_4,v+z_4}(y) \Big| \, dy \leqslant 2(a_3+a_4) \frac{|v-u|}{z_4} \, . \tag{3.8}$$

Finally, combining (3.5), (3.7) and (3.8), we have

$$\int_{\mathbb{R}_+} \big| k_G(x|u) - k_G(x|v) \big| \, dx \leqslant 2|v-u| \left[ \frac{a_1+a_2}{z_0} + \frac{a_3+a_4}{z_4} \right] .$$

□

Putting together Propositions 3.3 and 3.4, we have the following result.

PROPOSITION 3.5. — *Assume that $z_0 \leqslant z_4$, $a_1 + a_2 \geqslant 1$, $a_3 + a_4 \geqslant 1$, $a_1 + a_4 > 1$, and that $\lambda_G < 1$. Then for $x \in \mathbb{R}_+$ and $n \in \mathbb{N}$, we have*

$$\|P_G^n(x,\cdot) - \Pi_G(\cdot)\|_{TV} \leqslant \lambda_G^{n-1} \left[ \frac{z_4(a_2+a_3)}{a_1+a_4-1} + x \right] \left[ \frac{a_1+a_2}{z_0} + \frac{a_3+a_4}{z_4} \right] .$$

*Remark* 3.6. — It is important to reiterate that we assumed $z_0 \leqslant z_4$ only to simplify the exposition. An analogous result can be proven for the case where $z_0 > z_4$ by reversing the roles of these two quantities in the proof of Proposition 3.5.

Note that $\lambda_G$ is the expectation of a very simple function of two independent gamma random variables. Hence, it is a simple matter to use the classical Monte Carlo method to get an accurate estimate of $\lambda_G$. For the numerical example considered in the Introduction, a simple Monte Carlo experiment (strongly) suggests that $\lambda_G$ is slightly less than 0.62. Using 0.62 in place of 5/6 shows that 32 steps are actually enough. (Of course, we cannot conclude that $\lambda_G \leqslant 0.62$ based on Monte Carlo, no matter how large the sample is.)

## 4. A comparison to the bound of Steinsaltz

In this section, we compare the Wasserstein bound in Theorem 2.1 to an alternative bound from [9] (see also [6]). Assume that the Markov chain corresponding to $P$ can be represented as an iterated (random) function system; that is, assume that there exists a random function $f$ (with domain and range both equal $\mathsf{X}$) such that $f(x) \sim P(x,\cdot)$. (This is, of course, true of the chain analyzed in the previous section with $\mathsf{X} = \mathbb{R}_+$.) Assume further that the coupled Markov chains in Theorem 2.1 are based on the *same* iid sequence of random functions as at the beginning of Section 3. That is, $U_n = f_n \circ f_{n-1} \circ \cdots \circ f_1(u)$ and $V_n = f_n \circ f_{n-1} \circ \cdots \circ f_1(v)$, where $\{f_i\}_{i=1}^\infty$ is a sequence of iid random functions each with the same distribution as $f$. Finally, to keep things simple, assume that $g(x) = x$, so that $c = 1$. Under these assumptions, (2.1) is equivalent to

$$E\left[\frac{\big|f(v)-f(u)\big|}{|v-u|}\right] \leqslant \lambda \,, \tag{4.1}$$

for all $v \neq u$. Thus, for every $v \in \mathsf{X}$,

$$\limsup_{u\to v} E\left[\frac{\big|f(v)-f(u)\big|}{|v-u|}\right] \leqslant \lambda \,.$$

Then, assuming that this inequality still holds when limsup and integral are interchanged (as is the case for the chain analyzed in the previous section), we have

$$\sup_{v\in\mathsf{X}} E\left[\limsup_{u\to v} \frac{\big|f(v)-f(u)\big|}{|v-u|}\right] \leqslant \lambda \,.$$

Under these conditions, Theorem 2 in [9] is applicable (with $\phi(\cdot) \equiv 1$ and $r = \lambda$), and it follows that

$$d_W\big(P^n(x,\cdot),\Pi(\cdot)\big) \leqslant \frac{\lambda^n}{1-\lambda} E\big|f(x)-x\big| \,.$$

This is similar to the bound in Theorem 2.1, except that the term $E|Z-x|$ has been replaced by $\frac{E|f(x)-x|}{1-\lambda}$. To compare the bounds, suppose that $Z \sim \Pi$, and that $f$ and $Z$ are independent. Of course, $f(Z) \sim \Pi$ since $\Pi$ is stationary. Now, it follows from (4.1) that

$$E\big|f(Z)-f(x)\big| = E\Big[E\Big[\big|f(Z)-f(x)\big|\,\Big|\,Z\Big]\Big] \leqslant \lambda E\big|Z-x\big| \,.$$

Hence,

$$E\big|Z-x\big| = E\big|f(Z)-x\big| \leqslant E\big|f(Z)-f(x)\big| + E\big|f(x)-x\big| \leqslant \lambda E\big|Z-x\big| + E\big|f(x)-x\big| \,.$$

Consequently,

$$E|Z-x| \leqslant \frac{E|f(x)-x|}{1-\lambda} ,$$

which shows that, at least in the scenario we've described here, the bound in Theorem 2.1 is at least as good as that of [9]. Of course, in most practical applications, both $E|Z-x|$ and $E|f(x)-x|$ would be intractable integrals that would have to be either approximated analytically or estimated via Monte Carlo methods.

**Acknowledgment**. — The authors thank Persi Diaconis and an anonymous reviewer for helpful comments and suggestions. The first author was supported by NSF Grant DMS-11-06395, and the second by NSF Grant DMS-11-06084.

## Bibliography

[1] Batir (N.). — Inequalities for the gamma function. Archiv der Mathematik, 91, p. 554-563 (2008).

[2] Diaconis (P.), Khare (K.) and Saloff-Coste (L.). — Gibbs sampling, exponential families and orthogonal polynomials (with discussion). Statistical Science, 23, p. 151-200 (2008).

[3] Diaconis (P.), Khare (K.) and Saloff-Coste (L.). — Gibbs sampling, conjugate priors and coupling. Sankhyā: The Indian Journal of Statistics, 72-A, p. 136-169 (2009).

[4] Dudley (R. M.). — Real Analysis and Probability. Wadsworth, Belmont, CA (1989).

[5] Jovanovski (O.) and Madras (N.). — Convergence rates for hierarchical Gibbs samplers. Tech. rep., York University. ArXiv:1402.4733v2 (2014).

[6] Madras (N.) and Sezer (D.). — Quantitative bounds for Markov chain convergence: Wasserstein and total variation distance. Bernoulli, 16, p. 882-908 (2010).

[7] Meyn (S. P.) and Tweedie (R. L.). — Markov Chains and Stochastic Stability. Springer-Verlag, London (1993).

[8] Román (J. C.), Hobert (J. P.) and Presnell (B.). — On reparametrization and the Gibbs sampler. Statistics and Probability Letters, 91, p. 110-116 (2014).

[9] Steinsaltz (D.). — Locally contractive iterated function systems. The Annals of Probability, 27, p. 1952-1979 (1999).