Elect. Comm. in Probab. **7** (2002) 141–155

ELECTRONIC COMMUNICATIONS in PROBABILITY

FILL'S ALGORITHM FOR ABSOLUTELY CONTINUOUS STOCHASTICALLY MONOTONE KERNELS

MOTOYA MACHIDA
Department of Mathematics
Tennessee Technological University, Cookeville, TN 38505
email: machida@math.tntech.edu

submitted March 8, 2001 *Final version accepted* August 5, 2002

AMS 2000 Subject classification: Primary. 60J10, 68U20; Secondary. 60G40, 65C05, 65C10, 65C40.
Markov chain Monte Carlo, Fill's algorithm, perfect sampling, exact sampling, rejection sampling, stochastic monotonicity, partially ordered set, monotone coupling, absolutely continuous Markov kernel, regularity conditions.

Abstract
Fill, Machida, Murdoch, and Rosenthal (2000) presented their algorithm and its variants to extend the perfect sampling algorithm of Fill (1998) to chains on continuous state spaces. We consider their algorithm for absolutely continuous stochastically monotone kernels, and show the correctness of the algorithm under a set of certain regularity conditions. These conditions succeed in relaxing the previously known hypotheses sufficient for their algorithm to apply.

1 Introduction

1.1 An extension of Fill's algorithm

In Markov chain Monte Carlo (MCMC) a Markov kernel K is devised in an attempt to sample approximately from its stationary distribution π. Then the rate of convergence to stationarity could determine the Markov chain steps t to gain a desired approximation; however, such analysis is often found too difficult in practice. A perfect sampling algorithm, if applicable, not only eliminates the necessity of such analysis but also enables us to simulate exactly from π. The recent paper by Fill, Machida, Murdoch, and Rosenthal [3] laid down a framework for an extension of Fill's perfect sampling algorithm [1] to generic chains on general state spaces. We will call the extension of Fill's algorithm (and its variants) in [3] FMMR algorithms. The original form of this algorithm was proposed by Fill [1] for finite-state stochastically monotone chains, and was later extended by Thönnes [12] to the penetrable spheres mixture models, and by Møller and Schladitz [8] to general repulsive models. Their extensions were then considered in the context of FMMR algorithms (see Remark 7.4 in [3]).
In this paper we discuss the FMMR algorithm when a monotone case obtains, that is, when (a) a state space $\mathcal{X}$ is partially ordered with minimum element $\hat{0}$ and maximum element $\hat{1}$

(i.e., $\hat{0} \le x \le \hat{1}$ for every $x \in \mathcal{X}$), (b) the Markov kernel K is stochastically monotone (see Section 2.1 for a definition), and (c) there exists a collection $\{M_{x,y} : x \le y\}$ of upward Markov kernels satisfying for every $x \le y$,

$$K(y, \cdot) = \int K(x, dx') M_{x,y}(x', \cdot). \tag{1.1}$$

[We call a kernel L an *upward* kernel if the probability measure $L(x', \cdot)$ is, for every $x' \in \mathcal{X}$, supported on the set $\{y' : y' \ge x'\}$.] Given a collection of such upward kernels, the FMMR algorithm is a form of rejection sampling to simulate exactly from π.

Algorithm 1.1. In the first phase, we use the kernel $\tilde{K}$ (called the *time-reversed* kernel) satisfying $\pi(dx)K(x, dy) = \pi(dy)\tilde{K}(y, dx)$. Fix a positive integer t, and run a chain for t steps from the time-reversed kernel $\tilde{K}$, starting from $\boldsymbol{X}_t = \hat{0}$, and obtaining $\boldsymbol{X}_{t-1}, \ldots, \boldsymbol{X}_0$ in succession. In the second phase, generate a chain $(\boldsymbol{Y}_0, \ldots, \boldsymbol{Y}_t)$, starting from $\boldsymbol{Y}_0 = \hat{1}$, and obtaining $\boldsymbol{Y}_s$ inductively for $s = 1, \ldots, t$ from the distribution $M_{x_{s-1}, y_{s-1}}(x_s, \cdot)$ given $\boldsymbol{X}_{s-1} = x_{s-1}$, $\boldsymbol{Y}_{s-1} = y_{s-1}$ and $\boldsymbol{X}_s = x_s$. Finally check whether $\boldsymbol{Y}_t = \hat{0}$. If so, the value $\boldsymbol{X}_0$ is accepted as an observation from the stationary distribution π. If not, we start the routine again with an independent simulation (and possibly with a fresh choice of t).

The framework of FMMR algorithm indicates that Algorithm 1.1 runs correctly if the stationary distribution π is atomic at $\hat{0}$ [i.e., $\pi(\{\hat{0}\}) > 0$]; see Section 7 of [3]. In Section 1.2 we will present a simple example, and demonstrate that Algorithm 1.1 can fail without this "atomicity" condition, but runs correctly when different upward kernels are chosen. This illustrates our motive for studying the FMMR algorithm when a Markov kernel K is absolutely continuous, that is, when it has a kernel density k and a stationary distribution π (which is now a density). After making a short discussion on the issue of technical assumptions in Section 2.1, we propose a set of regularity conditions [namely, (R1)–(R3) in Section 2.2]. In simple terms, they are (R1) an everywhere positiveness and continuity at $\hat{0}$ for the stationary density π, (R2) an equicontinuity at $\hat{0}$ for the kernel density $k(x, \cdot)$ [see Remark 2.1(a)], and (R3) a (sort of) continuity property at $(\hat{0}, \{\hat{0}\})$ for the upward kernel $M_{x,y}$ [see Remark 2.1(b)]. Then our objective of the present paper is to show the correctness of Algorithm 1.1 under the regularity conditions, namely the following theorem.

Theorem 1.2. *Suppose that a monotone case obtains, and that a Markov kernel K is absolutely continuous. Then if the regularity conditions hold, then Algorithm 1.1 generates a sample exactly from stationary density π with success probability $k^t(\hat{1}, \hat{0})/\pi(\hat{0})$, where k^t denotes the t-step transition kernel density.*

In the rest of Section 2 (that is, Sections 2.3–2.5), we will present several key properties of the kernel density k and the upward kernel $M_{x,y}$, which leads to our proof of Theorem 1.2 in Section 3.1. In Section 3.2 we will discuss how we can construct a particular upward kernel having the continuity property in (R3). Finally in Section 3.3 we will introduce a quasi-monotone case toward possible extensions of Theorem 1.2, and discuss the applicability of FMMR algorithm for this quasi-monotone case. The present paper is what has become of the reference listed as [31] in [3].

1.2 A simple example

We present a simple example constructed on the state space $\mathcal{X} := [0, 1]$ with the usual Euclidean metric and with the usual linear order having the minimum element 0 and the maximum element 1. Let ν denote the Lebesgue measure on $[0,1]$, and let δ_x denote the Dirac measure at x. Then we can define a probability measure $\pi_\alpha(\cdot)$ on $[0,1]$ with parameter $\alpha \in [0,1/2)$ by

$$\pi_\alpha(\cdot) := \alpha\delta_0(\cdot) + \alpha\delta_1(\cdot) + (1-2\alpha)\nu(\cdot),$$

and a Markov kernel K (with parameter α) by

$$K(x,\cdot) := \begin{cases} \frac{3}{2}\pi_\alpha(\cdot \cap [0,1/2)) + \frac{1}{2}\pi_\alpha(\cdot \cap [1/2,1]) & \text{if } x < 1/2; \\ \frac{1}{2}\pi_\alpha(\cdot \cap [0,1/2)) + \frac{3}{2}\pi_\alpha(\cdot \cap [1/2,1]) & \text{if } x \geq 1/2. \end{cases}$$

It can be easily checked that the Markov kernel K is stochastically monotone, and satisfies $\pi_\alpha(dx)K(x,dy) = \pi_\alpha(dy)K(y,dx)$. Thus, K is a time-reversible Markov kernel with stationary distribution π_α.

Let U be a uniform random variable on $[0,1]$, and let

$$\phi(x,U) := \inf\{y \in \mathcal{X} : U \leq K(x,[0,y])\} \tag{1.2}$$

be the inverse probability transformation of $K(x,\cdot)$. Then we call ϕ a *transition rule for K* because of the property that $K(x,\cdot) = P(\phi(x,U) \in \cdot)$ for every $x \in \mathcal{X}$. Furthermore, ϕ is said to be *monotone* in the sense that $\phi(x,U) \leq \phi(y,U)$ whenever $x \leq y$. Therefore, we find that the conditional probability

$$M_{x,y}(x',\cdot) := P(\phi(y,U) \in \cdot \mid \phi(x,U) = x') \tag{1.3}$$

yields a desired upward kernel satisfying (1.1) for every $x \leq y$. Now that the monotone case has obtained, we can apply Algorithm 1.1 with the choice of $t = 1$. In the first phase we generate a sample $\boldsymbol{X}_0 = z$ from the distribution $\tilde{K}(0,\cdot)$ which equals $K(0,\cdot)$ by reversibility, and in the second phase we accept the value z with probability $M_{z,1}(0,\{0\})$.

We first assume that $\alpha > 0$. Then we can easily compute

$$M_{z,1}(0,\{0\}) = P(\phi(1,U) = 0 \mid \phi(z,U) = 0) = \begin{cases} 1/3 & \text{if } z < 1/2; \\ 1 & \text{if } z \geq 1/2, \end{cases} \tag{1.4}$$

and $P(\boldsymbol{X}_0 \in dz \mid \text{accept}) = K(0,dz)M_{z,1}(0,\{0\})/(1/2) = \pi_\alpha(dz)$. Thus, we have seen that Algorithm 1.1 works as desired when π_α is atomic at 0. We now suppose, however, that $\alpha = 0$. This is the case when K has the kernel density

$$k(x,y) := \begin{cases} 3/2 & \text{if } x,y < 1/2 \text{ or } x,y \geq 1/2; \\ 1/2 & \text{otherwise,} \end{cases}$$

and the uniform stationary density $\pi_0(\cdot) \equiv 1$ on $[0,1]$. Since $\phi(z,U) = 0$ in (1.2) uniquely implies $\phi(1,U) = 0$, Algorithm 1.1 always accepts the value z [i.e., $M_{z,1}(0,\{0\}) \equiv 1$] in the second phase, and incorrectly generates a sample from the density $k(0,\cdot)$.

Here we consider the multigamma coupler, introduced by Murdoch and Green [9], and construct different upward kernels particularly for $\alpha = 0$. Let $U \equiv (V,W)$ be a random vector with a Bernoulli random variable V having success probability $1/2$ and, independently, with a uniform random variable W on $[0,1]$. Then we can define a monotone transition rule ϕ_0 for K by

$$\phi_0(x,U) = \begin{cases} W & \text{if } V = 0; \\ W/2 & \text{if } V = 1 \text{ and } x < 1/2; \\ (1+W)/2 & \text{if } V = 1 \text{ and } x \geq 1/2. \end{cases} \tag{1.5}$$

Therefore, we can similarly obtain the upward kernel $M_{x,y}(x,\cdot)$ via (1.3) with ϕ_0 substituted for ϕ. Since $P(V=0 \mid \phi_0(z,U)=0)$ equals $1/3$ or 1 accordingly as $z<1/2$ or $z \ge 1/2$, we can compute the acceptance probability $M_{z,1}(0,\{0\})$ which coincides exactly with (1.4). This immediately implies that Algorithm 1.1 runs correctly even when π_0 is not atomic at 0.

2 Regularity conditions and their properties

2.1 The monotone case

In what follows we assume that the state space $\mathcal{X}$ is a Polish space (i.e., a complete separable metric space), and is equipped with a closed partial ordering [i.e., $\{(x,y) : x \le y\}$ is closed in $\mathcal{X} \times \mathcal{X}$]. The Polish space assumption provides a sufficient condition for existence of regular conditional probabilities (see, e.g., Dudley [4]), and was used implicitly in Section 1.2. In a typical application the state space can be chosen as a connected region of some Euclidean space $\mathbb{R}^n$ with the usual (coordinate-wise) partial order. Additional technical assumptions will be made about the monotone case throughout this paper, as described below.

Measure space. A subset G of $\mathcal{X}$ is called a *down-set* if it implies $y \in G$ whenever $y \le x$ for some $x \in G$, and $\mathcal{D}$ will denote the collection of non-empty open down-sets. Then we assume that $\mathcal{D}$ is a base of neighborhoods of $\hat{0}$, that is, that

$$\{G' \in \mathcal{D} : G' \subseteq G\} \neq \emptyset \quad \text{for every neighborhood } G \text{ of } \hat{0}. \tag{2.1}$$

Let $B(\hat{0},r)$ denote the open ball of radius r from $\hat{0}$. Then $\mathcal{C} := \{B(\hat{0},1/n) : n \in \mathbb{N}\}$ is a countable base for neighborhoods of $\hat{0}$. In the usual (coordinate-wise) partial order on some Euclidean space, every open ball $B(\hat{0},r)$ is a down-set, and therefore, (2.1) holds. Although an open ball $B(\hat{0},r)$ is not necessarily a down-set in general, (2.1) still holds if $\mathcal{X}$ is compact (see Theorem 5 on page 48 of [10]).

Let $(\mathcal{X},\mathcal{B},\nu)$ be a measure space with Borel σ-field $\mathcal{B}$ and σ-finite measure ν. Then we assume that

$$\nu(G) > 0 \quad \text{for every neighborhood } G \text{ of } \hat{0}. \tag{2.2}$$

We note that (2.2) is equivalently characterized by every nonnegative measurable function g for which $\int g(x)\,\nu(dx) > 0$ whenever g is strictly positive and continuous at $\hat{0}$. The measure ν will be used as a "reference" measure to introduce density functions (i.e., Radon–Nykodym derivatives with respect to ν). For this reason we will simply write dz for $\nu(dz)$ as a shorthand notation. Lebesgue measure is the usual choice for reference measure ν when $\mathcal{X}$ is a connected region of some Euclidean space. But note that (2.2) implies $\nu(\{\hat{0}\}) > 0$ if the singleton $\{\hat{0}\}$ itself is a neighborhood of $\hat{0}$, and that this is clearly the case when $\mathcal{X}$ is discrete, or $\hat{0}$ is an isolated point in $\mathcal{X}$.

Monotone coupling. The notion of stochastic monotonicity is closely related to Nachbin–Strassen theorem on stochastic ordering, introduced by Kamae, Krengel, and O'Brien [6]. Among several equivalent definitions, stochastic ordering can be defined in terms of the collection $\mathcal{D}$ of non-empty open down-sets. Let P_1 and P_2 be probability measures on $(\mathcal{X},\mathcal{B})$. Then P_1 is said to be *stochastically smaller* than P_2, denoted by $P_1 \preceq P_2$, if $P_1(G) \ge P_2(G)$ for all $G \in \mathcal{D}$. The Nachbin–Strassen theorem (Theorem 1 in [6]) shows that $P_1 \preceq P_2$ if and only if there exists an upward kernel L satisfying $P_2(\cdot) = \int P_1(dx)L(x,\cdot)$. (See [6] for other characterizations of stochastic ordering.) A Markov kernel K is said to be *stochastically monotone* if $K(x,\cdot) \preceq K(y,\cdot)$ whenever $x \le y$. By the Nachbin–Strassen theorem the stochastic monotonicity of K is equivalent to existence of a collection $\{M_{x,y} : x \le y\}$ of upward kernels satisfying (1.1).

Let $\mathcal{V}$ denote the closed subset $\{(x,y) : x \le y\}$ of $\mathcal{X} \times \mathcal{X}$. Then we will assume that the function $M_{x,y}(x', A)$ of (x, y, x') is jointly measurable on $\mathcal{V} \times \mathcal{X}$ for every $A \in \mathcal{B}$, and call $M_{x,y}(x', \cdot)$ a *monotone coupling*. As we have demonstrated in Section 1.2, it is often the case that a monotone transition rule ϕ exists and that a monotone coupling $M_{x,y}(x', \cdot)$ can be constructed via (1.3). We note, however, that such monotone transition rule ϕ does not necessarily exist for stochastically monotone kernel K when $\mathcal{X}$ is not a linearly ordered state space. (For the precise statement in the case of a finite state space, consult [2].)

2.2 Regularity conditions

We define the essential supremum for a measurable function f on a measurable set G by

$$\operatorname*{ess.sup}_{x \in G} |f(x)| := \inf \{\alpha \ge 0 : \nu(\{x \in G : |f(x)| > \alpha\}) = 0\}.$$

We will say that a sequence $\{G_n\}_{n=1}^{\infty}$ of neighborhoods of $\hat{0}$ is *shrinking* if there is a sequence $\{B(\hat{0}, r_n)\}_{n=1}^{\infty}$ of open balls with $\lim_{n\to\infty} r_n = 0$ such that $G_n \subseteq B(\hat{0}, r_n)$. Now assume that the Markov kernel K has a kernel density k and a stationary density π. Then we can introduce the following conditions (R1)–(R3) for the monotone case, and call them collectively *regularity conditions*:

(R1) The stationary density π is continuous at $\hat{0}$, and satisfies $\pi(x) > 0$ for all $x \in \mathcal{X}$.

(R2) The kernel density k satisfies for every sequence $\{y_n\}_{n=1}^{\infty}$ converging to $\hat{0}$,

$$\lim_{n\to\infty} \sup_{x\in\mathcal{X}} |k(x, y_n) - k(x, \hat{0})| = 0.$$

(R3) If $k(x, \hat{0}) > 0$, then each upward kernel $M_{x,y}$ satisfies for every shrinking sequence $\{G_n\}_{n=1}^{\infty}$ of neighborhoods of $\hat{0}$,

$$\lim_{n\to\infty} \operatorname*{ess.sup}_{x' \in G_n} |M_{x,y}(x', G_n) - M_{x,y}(\hat{0}, \{\hat{0}\})| = 0.$$

Recall the example discussed in Section 1.2. Here we examine the regularity conditions when the Markov kernel K is absolutely continuous, that is, when $\alpha = 0$. Clearly the stationary density π_0 and the kernel density k satisfy (R1) and (R2), respectively. But the upward kernel (1.3) with transition rule (1.2) does not satisfy (R3). Indeed one can show that the limit in (R3) becomes unity by choosing $x < 1/2 \le y$ and the shrinking sequence $\{G_n\}_{n=1}^{\infty}$ with $G_n = [0, 1/n)$. In Section 1.2 we then obtained a successful monotone coupling by using another transition rule (1.5). We will see in Remark 2.1(b) that (R3) holds for this monotone coupling.

Remark 2.1. (a) Since $\mathcal{X}$ is a metric space, (R2) holds if and only if for any $\varepsilon > 0$ there exists some neighborhood G of $\hat{0}$ such that

$$|k(x,y) - k(x, \hat{0})| < \varepsilon \quad \text{for all } x \in \mathcal{X} \text{ and all } y \in G. \tag{2.3}$$

In the sense of (2.3) we will say that the functions $f(x, \cdot)$, $x \in \mathcal{X}$, are equicontinuous at $\hat{0}$.

(b) It suffices for (R3) to show that for any $\varepsilon > 0$ there exists some neighborhood G of $\hat{0}$ such that

$$|M_{x,y}(x', G') - M_{x,y}(\hat{0}, \{\hat{0}\})| \le \varepsilon \quad \text{for almost everywhere (a.e.) } x' \in G', \tag{2.4}$$

for any neighborhood G' of $\hat{0}$ satisfying $G' \subseteq G$. For example, consider the upward kernel (1.3) constructed from the "multigamma" transition rule (1.5) in Section 1.2. Let G' be any neighborhood G' of $\hat{0}$ such that $G' \subseteq G := [0, 1/3)$. Then, for every $x' \in G'$, we can calculate $M_{x,y}(x', G') = M_{x,y}(\hat{0}, \{\hat{0}\}) = 1/3$ if $x < 1/2 \le y$; otherwise, $M_{x,y}(x', G') = M_{x,y}(\hat{0}, \{\hat{0}\}) = 1$. Therefore, (2.4) holds for every $x \le y$.
(c) We can particularly choose a shrinking sequence $\{G_n\}_{n=1}^{\infty}$ of non-empty open down-sets. For each $n = 1, 2, \ldots$, by (2.1) we can find some $G_n \in \mathcal{D}$ such that $G_n \subseteq B(\hat{0}, 1/n)$. Then the resulting sequence $\{G_n\}_{n=1}^{\infty}$ is clearly shrinking.

2.3 Kernel density and ergodicity

A kernel density k^t exists for the t-step transition kernel K^t for $t = 1, 2, \ldots$, and is constructed recursively via

$$k^t(x, y) = \int k^{t-1}(x, z) k(z, y)\, dz \qquad \text{for } t = 2, 3, \ldots \tag{2.5}$$

where $k^1 \equiv k$. From the integral representation (2.5) one easily conclude that the equicontinuity property in (R2) is maintained for the t-step kernel density k^t. Recalling Remark 2.1(a), for any $\varepsilon > 0$ we can find some neighborhood G of $\hat{0}$ satisfying (2.3), and apply it to obtain

$$|k^t(x, y) - k^t(x, \hat{0})| \le \int k^{t-1}(x, z) |k(z, y) - k(z, \hat{0})|\, dz \le \varepsilon \tag{2.6}$$

for any $y \in G$, and for every $x \in \mathcal{X}$ and every $t = 1, 2, \ldots$. Thus, the choice of neighborhood G of $\hat{0}$ does not depend on t. From this observation in (2.6) we derive the following theorem in connection with ergodicity of the Markov kernel K.

Theorem 2.2. *For every $x \in \mathcal{X}$ we have $\lim_{t \to \infty} k^t(x, \hat{0}) = \pi(\hat{0})$.*

Proof. By (R1) and (2.6), for any $\varepsilon > 0$ we can find a neighborhood G of $\hat{0}$ such that

$$1 - \varepsilon < \frac{\pi(z)}{\pi(\hat{0})} < 1 + \varepsilon \quad \text{and} \quad \frac{k^t(x, \hat{0})}{\pi(\hat{0})} - \varepsilon < \frac{k^t(x, z)}{\pi(\hat{0})} < \frac{k^t(x, \hat{0})}{\pi(\hat{0})} + \varepsilon$$

for all $z \in G$ and all $t \ge 1$. By combining these two inequalities, we can obtain

$$\frac{k^t(x, \hat{0}) - \varepsilon\pi(\hat{0})}{(1 + \varepsilon)\pi(\hat{0})} \pi(z) < k^t(x, z) < \frac{k^t(x, \hat{0}) + \varepsilon\pi(\hat{0})}{(1 - \varepsilon)\pi(\hat{0})} \pi(z).$$

By integrating the above inequality over the neighborhood G, we derive

$$\frac{K^t(x, G)}{\pi(G)} (1 - \varepsilon) - \varepsilon \le \frac{k^t(x, \hat{0})}{\pi(\hat{0})} \le \frac{K^t(x, G)}{\pi(G)} (1 + \varepsilon) + \varepsilon \quad \text{for all } t \ge 1,$$

where we are using the symbol π for both density and distribution. By ergodicity $K^t(x, G)$ converges to $\pi(G)$ as $t \to \infty$, and

$$1 - 2\varepsilon \le \lim_{t \to \infty} \frac{k^t(x, \hat{0})}{\pi(\hat{0})} \le 1 + 2\varepsilon.$$

This completes the proof since $\varepsilon > 0$ is arbitrary. □

2.4 Pointwise derivative and upward kernel

Since $k^t(x,\cdot)$ is continuous at $\hat{0}$, we can consider the pointwise derivative of $K^t(y,\cdot)$ with respect to $K^t(x,\cdot)$ at $\hat{0}$ (see, e.g., [5]). Assuming $k^t(x,\hat{0}) > 0$, one can easily show that for any shrinking sequence $\{G_n\}_{n=1}^{\infty}$ of neighborhoods of $\hat{0}$,

$$\lim_{n\to\infty} \frac{K^t(y,G_n)}{K^t(x,G_n)} = \frac{k^t(y,\hat{0})}{k^t(x,\hat{0})}. \tag{2.7}$$

An immediate application of (2.7) comes from the fact that stochastic monotonicity for the kernel K is preserved for the t-step transition kernel K^t (see, e.g., [7]). Recall in Remark 2.1(c) that we can choose a shrinking sequence $\{G_n\}_{n=1}^{\infty}$ of non-empty open down-sets. Since $K^t(x,G_n) \geq K^t(y,G_n)$ for every $n \geq 1$, by (2.7) we conclude the following corollary.

Corollary 2.3. *If $x \leq y$, then $k^t(x,\hat{0}) \geq k^t(y,\hat{0})$.*

In the next proposition we will see that the pointwise derivative is also related to the upward kernel. Then, by combining Proposition 2.4 and (2.7), one can see that for each $(x,y) \in \mathcal{V}$,

$$M_{x,y}(\hat{0},\{\hat{0}\}) = \frac{k(y,\hat{0})}{k(x,\hat{0})} \tag{2.8}$$

whenever $k(x,\hat{0}) > 0$.

Proposition 2.4. *Assuming $k(x,\hat{0}) > 0$, we have for any shrinking sequence $\{G_n\}_{n=1}^{\infty}$ of non-empty down-sets,*

$$\lim_{n\to\infty} \frac{K(y,G_n)}{K(x,G_n)} = M_{x,y}(\hat{0},\{\hat{0}\}) \quad \textit{for } (x,y) \in \mathcal{V}.$$

Proof. Since $M_{x,y}$ is upward and G_n is down-set, we have $M_{x,y}(x',G_n) = 0$ for every $x' \notin G_n$. By (1.1) we obtain

$$K(y,G_n) = \int_{G_n} M_{x,y}(x',G_n)\,k(x,x')\,dx'.$$

Therefore, we can derive

$$\begin{aligned}
&\left|\frac{K(y,G_n)}{K(x,G_n)} - M_{x,y}(\hat{0},\{\hat{0}\})\right| \\
&\quad = \frac{1}{K(x,G_n)} \left|\int_{G_n} M_{x,y}(x',G_n)\,k(x,x')\,dx' - M_{x,y}(\hat{0},\{\hat{0}\}) \int_{G_n} k(x,x')\,dx'\right| \\
&\quad \leq \frac{1}{K(x,G_n)} \int_{G_n} \left|M_{x,y}(x',G_n) - M_{x,y}(\hat{0},\{\hat{0}\})\right| k(x,x')\,dx' \\
&\quad \leq \operatorname*{ess.sup}_{x'\in G_n} \left|M_{x,y}(x',G_n) - M_{x,y}(\hat{0},\{\hat{0}\})\right|.
\end{aligned}$$

Then the limit of the essential supremum must be zero by (R3). □

2.5 The t-fold monotone coupling

In this subsection we discuss the monotone coupling and introduce its t-fold extension under the regularity conditions. For each fixed $(x, y) \in \mathcal{V}$, we first define a measure on $\mathcal{V}$ by

$$\hat{K}((x,y), dx' \times dy') := M_{x,y}(x', dy') k(x, x') dx'.$$

From a simple constructive argument of Lebesgue integration, one can see that (1.1) equivalently implies that

$$\int g(y')\, k(y, y')\, dy' = \iint g(y')\, \hat{K}((x,y), dx' \times dy') \tag{2.9}$$

for every measurable function g on $\mathcal{X}$, if it is integrable. By the upwardness and the joint measurability of $M_{x,y}(x', \cdot)$, $\hat{K}$ is actually a Markov kernel on the state space $\mathcal{V}$. Then we can establish the following lemma.

Lemma 2.5. *Let $g(u, v, x')$ be a jointly measurable function on $\mathcal{V} \times \mathcal{X}$. Then*

$$\hat{g}(x, y, x') = \iint g(u, v, x')\, \hat{K}((x,y), du \times dv)$$

is jointly measurable on $\mathcal{V} \times \mathcal{X}$ if it is integrable.

Proof. We can prove it by monotone class argument (see, e.g., Dudley [4]). First let $g(u, v, x') = 1_{A \times B}(u, v, x')$ be an indicator function on $A \times B$ with measurable sets $A \subseteq \mathcal{V}$ and $B \subseteq \mathcal{X}$. Then clearly $\hat{g}(x, y, x') = \hat{K}((x,y), A) 1_B(x')$ is jointly measurable. We can then pass to finite disjoint unions of them, and then to any measurable sets on $\mathcal{V} \times \mathcal{X}$ by monotone class theorem. Now g can be replaced by simple functions, nonnegative and general jointly measurable functions. □

Now we define the collection $\{M^t{}_{x,y} : (x, y) \in \mathcal{V}\}$ of t-fold upward kernels as follows. We set $M^1{}_{x,y}(x', \cdot) := M_{x,y}(x', \cdot)$, and for $t \geq 2$, we can construct $M^t{}_{x,y}(x', \cdot)$ recursively via

$$M^t{}_{x,y}(x', \cdot) := \frac{1}{k^t(x, x')} \iint M^{t-1}{}_{u,v}(x', \cdot) k^{t-1}(u, x') \hat{K}((x,y), du \times dv) \tag{2.10}$$

if $k^t(x, x') > 0$; otherwise, we set $M^t{}_{x,y}(x', \cdot)$ to be an arbitrary upward kernel, say $M^t{}_{x,y}(x', \cdot) := \delta_{x'}(\cdot)$. Then it is an immediate corollary of Lemma 2.5 that the function $M^t{}_{x,y}(x', \cdot)$ of (x, y, x') is jointly measurable. Furthermore, it satisfies

$$K^t(y, \cdot) = \int M^t{}_{x,y}(x', \cdot) k^t(x, x')\, dx' \quad \text{for every } (x, y) \in \mathcal{V}, \tag{2.11}$$

which is verified by the following straightforward induction argument: By applying Fubini's theorem and then (2.9), one can show that

$$\begin{aligned}\int M^t{}_{x,y}(x', \cdot) k^t(x, x')\, dx' &= \iint \left[\int M^{t-1}{}_{u,v}(x', \cdot) k^{t-1}(u, x')\, dx' \right] \hat{K}((x,y), du \times dv) \\ &= \iint K^{t-1}(v, \cdot)\, \hat{K}((x,y), du \times dv) = \int K^{t-1}(v, \cdot)\, k(y, v)\, dv = K^t(y, \cdot).\end{aligned}$$

The next proposition is the result of another straightforward induction argument, and will be the most important one in the proof of Theorem 1.2.

Proposition 2.6. *If $x \le y$ and $k^t(x,\hat{0}) > 0$, then we have*

$$M^t{}_{x,y}(\hat{0},\{\hat{0}\}) = \frac{k^t(y,\hat{0})}{k^t(x,\hat{0})}. \tag{2.12}$$

Proof. Note that (2.8) is a special case of (2.12) for $t=1$. For $t \ge 2$, by applying the induction hypothesis and then (2.9), we derive

$$\begin{aligned} M^t{}_{x,y}(\hat{0},\{\hat{0}\}) &= \frac{1}{k^t(x,\hat{0})}\iint M^{t-1}{}_{u,v}(\hat{0},\{\hat{0}\})k^{t-1}(u,\hat{0})\hat{K}((x,y),du\times dv) \\ &= \frac{1}{k^t(x,\hat{0})}\iint k^{t-1}(v,\hat{0})\,\hat{K}((x,y),du\times dv) \\ &= \frac{1}{k^t(x,\hat{0})}\iint k^{t-1}(v,\hat{0})k(y,v)\,dv = \frac{k^t(y,\hat{0})}{k^t(x,\hat{0})}. \end{aligned}$$

□

In the sense of (2.11), we will call $M^t{}_{x,y}(x',\cdot)$ a *t-fold monotone coupling.* Then the last theorem of this subsection shows that the continuity property of (R3) is maintained for the t-fold extension of $M_{x,y}(x',\cdot)$. This in turn can be used to provide an alternative proof of Proposition 2.6 which will be exactly the same as that of Proposition 2.4.

Theorem 2.7. *For every $t \ge 1$ and for every shrinking sequence $\{G_n\}_{n=1}^{\infty}$ of neighborhoods of $\hat{0}$, we have*

$$\lim_{n\to\infty} \operatorname*{ess.sup}_{x'\in G_n} |M^t{}_{x,y}(x',G_n) - M^t{}_{x,y}(\hat{0},\{\hat{0}\})| = 0$$

if $k^t(x,\hat{0}) > 0$.

Proof. For each $n \ge 1$ define

$$B_n := \{x' \in G_n : |M_{x,y}(x',G_n) - M_{x,y}(\hat{0},\{\hat{0}\})| > a_n\},$$

where $a_n := \text{ess.sup}_{x'\in G_n} |M_{x,y}(x',G_n) - M_{x,y}(\hat{0},\{\hat{0}\})|$. By definition of the essential supremum, we have $\nu(B_n) = 0$ for every $n \ge 1$, and therefore, $\nu\left(\bigcup_{n=1}^{\infty} B_n\right) = 0$. Let $G'_n := G_n \setminus \left(\bigcup_{n=1}^{\infty} B_n\right)$. Then if it can be shown that

$$\lim_{n\to\infty} \sup_{x'\in G'_n} M^t{}_{x,y}(x',G_n) = \lim_{n\to\infty} \inf_{x'\in G'_n} M^t{}_{x,y}(x',G_n) = M^t{}_{x,y}(\hat{0},\{\hat{0}\}), \tag{2.13}$$

then the proof will be completed.
We will show (2.13) by induction. By (R3) we can see that (2.13) clearly holds for $t=1$. Since $M^{t-1}{}_{u,v}(x',G_n) \le 1$ and $k^{t-1}(u,\cdot)$ is equicontinuous at $\hat{0}$, we can find some integer n_0 and positive value $c > 0$ satisfying

$$\sup_{x'\in G'_n} \left[M^{t-1}{}_{u,v}(x',G_n)k^{t-1}(u,x')\right] \le c$$

for all $(u, v) \in \mathcal{V}$ and all $n \geq n_0$. By applying the induction hypothesis and then dominated convergence theorem, we derive

$$\begin{aligned}
&\lim_{n\to\infty} \sup_{x'\in G'_n} M^t{}_{x,y}(x', G_n) \\
&\quad \leq \lim_{n\to\infty} \left[\sup_{x'\in G'_n} \frac{1}{k^t(x,x')} \right] \left[\iint \sup_{x'\in G'_n} \left[M^{t-1}{}_{u,v}(x', G_n) k^{t-1}(u, x') \right] \hat{K}((x,y), du \times dv) \right] \\
&\quad = \frac{1}{k^t(x,\hat{0})} \iint M^{t-1}{}_{u,v}(\hat{0}, \{\hat{0}\})\, k^{t-1}(u,\hat{0}) \hat{K}((x,y), du \times dv) = M^t{}_{x,y}(\hat{0}, \{\hat{0}\}),
\end{aligned}$$

and

$$\begin{aligned}
&\lim_{n\to\infty} \inf_{x'\in G'_n} M^t{}_{x,y}(x', G_n) \\
&\quad \geq \lim_{n\to\infty} \left[\inf_{x'\in G'_n} \frac{1}{k^t(x,x')} \right] \left[\iint \inf_{x'\in G'_n} \left[M^{t-1}{}_{u,v}(x', G_n) k^{t-1}(u, x') \right] \hat{K}((x,y), du \times dv) \right] \\
&\quad = \frac{1}{k^t(x,\hat{0})} \iint M^{t-1}{}_{u,v}(\hat{0}, \{\hat{0}\})\, k^{t-1}(u,\hat{0}) \hat{K}((x,y), du \times dv) = M^t{}_{x,y}(\hat{0}, \{\hat{0}\}).
\end{aligned}$$

They together implies (2.13). $\square$

3 FMMR algorithm for absolutely continuous kernels

3.1 Proof of Theorem 1.2

We now complete the proof of Theorem 1.2 by introducing a rejection sampling method (see, e.g., Chapter 10 of [11]), which essentially parallels Section 7.3 of [1]. Under the regularity conditions we can form a time-reversal kernel density $\tilde{k}$ by

$$\tilde{k}(y, x) = \frac{\pi(x) k(x,y)}{\pi(y)}, \tag{3.1}$$

and obtain the t-step time-reversed kernel density $\tilde{k}^t$ via (2.5) with $\tilde{k}$ substituted for k. Suppose that there exists a constant c satisfying

$$\frac{\pi(z)}{\tilde{k}^t(\hat{0}, z)} \leq c \quad \text{for all } z \in \mathcal{X}. \tag{3.2}$$

Then one can construct an algorithm composed of the following two steps:

Step 1. Generate a sample value $\boldsymbol{X}_0 = z$ from the density $\tilde{k}^t(\hat{0}, \cdot)$.

Step 2. Accept the value $\boldsymbol{X}_0 = z$ with acceptance probability $c^{-1}\pi(z) \Big/ \tilde{k}^t(\hat{0}, z)$; otherwise, return to *Step 1.*

According to the rejection sampling principle, this algorithm generates a sample exactly from the density π, and will yield an accepted value every iteration with success probability c^{-1}. And it will perform best when the least upper bound c^* in (3.2) is chosen. By applying (2.5) with (3.1) inductively, we obtain $\pi(x)k^t(x,y) \equiv \pi(y)\tilde{k}^t(y,x)$. Then, together with Corollary 2.3 we derive the least upper bound

$$\frac{\pi(z)}{\tilde{k}^t(\hat{0},z)} = \frac{\pi(\hat{0})}{k^t(z,\hat{0})} \leq c^* = \frac{\pi(\hat{0})}{k^t(\hat{1},\hat{0})}, \tag{3.3}$$

and the acceptance probability

$$\frac{k^t(\hat{1},\hat{0})}{\pi(\hat{0})} \cdot \frac{\pi(z)}{\tilde{k}^t(\hat{0},z)} = \frac{k^t(\hat{1},\hat{0})}{k^t(z,\hat{0})}. \tag{3.4}$$

By Theorem 2.2 and (R1), note that we can choose sufficiently large Markov chain steps t in order for (3.3) and (3.4) to have $k^t(\hat{1},\hat{0}) > 0$.
Now we will see that Algorithm 1.1 is designed precisely to realize the above-mentioned algorithm. Clearly the first phase of Algorithm 1.1 is Step 1. The acceptance probability of Algorithm 1.1 can be expressed in the closed form

$$\int \cdots \int \frac{\tilde{k}(\hat{0},x_{t-1})\cdots\tilde{k}(x_1,z)}{\tilde{k}^t(\hat{0},z)} M_{z,\hat{1}}(x_1,dy_1)\cdots M_{x_{t-1},y_{t-1}}(\hat{0},\{\hat{0}\})\,dx_1\cdots dx_{t-1}.$$

By recursively applying (2.10), the t-fold upward kernel ${M^t}_{z,\hat{1}}(\hat{0},\{\hat{0}\})$ has also the closed from

$$\int \cdots \int M_{z,\hat{1}}(x_1,dy_1)\cdots M_{x_{t-1},y_{t-1}}(\hat{0},\{\hat{0}\})\,\frac{k(z,x_1)\cdots k(x_{t-1},\hat{0})}{k^t(z,\hat{0})}\,dx_1\cdots dx_{t-1}.$$

Since

$$\frac{\tilde{k}(\hat{0},x_{t-1})\cdots\tilde{k}(x_1,z)}{\tilde{k}^t(\hat{0},z)} = \frac{k(z,x_1)\cdots k(x_{t-1},\hat{0})}{k^t(z,\hat{0})},$$

one can see that the acceptance probability in Algorithm 1.1 is exactly equal to the probability ${M^t}_{z,\hat{1}}(\hat{0},\{\hat{0}\})$. Furthermore, by Proposition 2.6 the second phase of Algorithm 1.1 works precisely as Step 2, and generates the same acceptance probability (3.4). Therefore, we have shown the correctness of Algorithm 1.1.

3.2 Gamma coupling

Let $\mathcal{W}$ be a measurable set on $\mathcal{X} \times \mathcal{X}$. By $\mathcal{W}_x := \{y : (x,y) \in \mathcal{W}\}$ we denote the section determined by the first coordinate x. In what follows we will assume that $\mathcal{W}_x$ is nonempty for every $x \in \mathcal{X}$. Then a collection $\{M_{x,y} : (x,y) \in \mathcal{W}\}$ of Markov kernels is called a coupling, if (i) $M_{x,y}(x',\cdot)$ is supported on the section $\mathcal{W}_{x'}$ for every $(x,y,x') \in \mathcal{W} \times \mathcal{X}$, and (ii) the function $M_{x,y}(x',A)$ of (x,y,x') is jointly measurable on $\mathcal{W} \times \mathcal{X}$ for each $A \in \mathcal{B}$. We simply write $M_{x,y}(x',\cdot)$ for a coupling if the underlying set $\mathcal{W}$ has been clearly specified. Furthermore, given a Markov kernel K, we call $M_{x,y}(x',\cdot)$ a *coupling for* K if it is a coupling satisfying (1.1) for every $(x,y) \in \mathcal{W}$. We note that a monotone coupling is a special case of coupling for K where the set $\mathcal{W}$ is particularly chosen to be the closed set $\mathcal{V} = \{(x,y) : x \leq y\}$. One can generate a bivariate chain $(X_t,Y_t)_{t=0,1,\ldots}$ on the state space $\mathcal{W}$ from the bivariate Markov kernel

$$\hat{K}((x,y),dx' \times dy') := M_{x,y}(x',dy')K(x,dx'), \tag{3.5}$$

and the initial distribution (X_0,Y_0). Then the marginal chain $(X_t)_{t=0,1,\ldots}$ is clearly a Markov chain generated by K, and (1.1) becomes a necessary and sufficient condition for the marginal

chain $(Y_t)_{t=0,1,\ldots}$ to be Markovian, having the same Markov kernel K. When $\mathcal{W} = \mathcal{X} \times \mathcal{X}$, we can always construct the "trivial" coupling $M_{x,y}(x',\cdot) := K(y,\cdot)$ for a Markov kernel K, by which one can run independent marginal chains $(X_t)_{t=0,1,\ldots}$ and $(Y_t)_{t=0,1,\ldots}$.
In the rest of this subsection we further assume that $\mathcal{W}$ contains the diagonal set $\Delta := \{(x,x) : x \in \mathcal{X}\}$, and that the Markov kernel K has a kernel density k. Given a coupling $L_{x,y}(x',\cdot)$, we can construct another coupling $M_{x,y}(x',\cdot)$ by

$$M_{x,y}(x',\cdot) := \begin{cases} \delta_{x'}(\cdot) & \text{if } k(x,x') \le k(y,x'); \\ \frac{k(y,x')}{k(x,x')}\,\delta_{x'}(\cdot) + \frac{k(x,x')-k(y,x')}{k(x,x')}\,L_{x,y}(x',\cdot) & \text{if } k(x,x') > k(y,x'). \end{cases} \tag{3.6}$$

Let $(x,y) \in \mathcal{W}$ be fixed, and let $\gamma_{x,y} := \int k(x,z) \wedge k(y,z)\,dz$ where $a \wedge b := \min(a,b)$. Assuming $\gamma_{x,y} < 1$, we can introduce the probability measures

$$Q_{(x,y)}(dz) := \frac{1}{1-\gamma_{x,y}}\,[k(x,z) - k(x,z) \wedge k(y,z)]\,dz$$

and

$$Q^{(x,y)}(dz) := \frac{1}{1-\gamma_{x,y}}\,[k(y,z) - k(x,z) \wedge k(y,z)]\,dz.$$

If $L_{x,y}(x',\cdot)$ satisfies for every $(x,y) \in \mathcal{W}$,

$$Q^{(x,y)}(A) = \int L_{x,y}(x',A)\,Q_{(x,y)}(dx') \quad \text{for each } A \in \mathcal{B}, \tag{3.7}$$

then we will see in the next lemma that (3.6) is a coupling for K, and call it a *γ-coupling*. Note that if $\gamma_{x,y} = 1$ then $k(x,x') = k(y,x')$ and $M_{x,y}(x',\cdot) = \delta_{x'}(\cdot)$ for a.e. $x' \in \mathcal{X}$, which trivially satisfies (1.1).

Lemma 3.1. *The coupling $M_{x,y}(x',\cdot)$ in (3.6) satisfies (1.1) if and only if $L_{x,y}(x',\cdot)$ satisfies (3.7).*

Proof. Define $B := \{x' : k(x,x') \le k(y,x')\}$. By observing that $Q_{(x,y)}(B) = 0$ and $Q^{(x,y)}(B^c) = 0$, we can obtain

$$\begin{aligned} \int M_{x,y}(x',A)\,k(x,x')\,dx' &= \int_B \delta_{x'}(A)\,k(x,x')\,dx' \\ &\quad + \int_{B^c} \left[\frac{k(y,x')}{k(x,x')}\,\delta_{x'}(A) + \frac{k(x,x')-k(y,x')}{k(x,x')}\,L_{x,y}(x',A)\right] k(x,x')\,dx' \\ &= K(y,A) - (1-\gamma_{x,y})\left[Q^{(x,y)}(A) - \int L_{x,y}(x',A)\,Q_{(x,y)}(dx')\right], \end{aligned}$$

which equals $K(y,A)$ if and only if (3.7) holds. □

The notion of γ-coupling appeared at several different places (see the notes in Chapter 1 of [7]), and was introduced by Lindvall [7] for a coupling of two probability measures P_1 and P_2 to achieve a maximal coupling. This particular coupling was a bivariate probability measure Q on a product space whose marginals are consistent with P_1 and P_2, and its mass $Q(\Delta)$ on the diagonal set Δ achieves its maximum $\left(1 - \frac{1}{2}\|P_1 - P_2\|\right)$, where $\|\cdot\|$ denotes the total variation norm. By employing the same argument as in [7], one can show that $\gamma_{x,y} = \left(1 - \frac{1}{2}\|K(x,\cdot) - K(y,\cdot)\|\right)$, and that our γ-coupling in (3.6) attains $\hat{K}((x,y),\Delta) = \gamma_{x,y}$. But in the next proposition we justify our notion of the γ-coupling in connection with the regularity conditions.

Proposition 3.2. *If the kernel density k satisfies (R2), then (R3) holds for a γ-coupling $M_{x,y}(x',\cdot)$ of the form (3.6).*

Proof. Let $\varepsilon > 0$ be arbitrary. By Remark 2.1(b), (R3) holds if we can find a neighborhood G of $\hat{0}$ so that (2.4) holds for any neighborhood G' of $\hat{0}$ satisfying $G' \subseteq G$. (i) Suppose that $0 < k(x,\hat{0}) < k(y,\hat{0})$. Then by (R2) we can find some neighborhood G of $\hat{0}$ such that $0 < k(x,x') < k(y,x')$ for all $x' \in G$. Thus, we obtain

$$|M_{x,y}(x',G') - M_{x,y}(\hat{0},\{\hat{0}\})| = |\delta_{x'}(G') - \delta_{\hat{0}}(\{\hat{0}\})| = 0$$

for every $x' \in G'$ and for every measurable subset G' of G, as desired. (ii) Suppose that $0 < k(x,\hat{0}) = k(y,\hat{0})$. Similarly we can find some neighborhood G of $\hat{0}$ such that $0 < k(x,x')$ and $|k(x,x') - k(y,x')|/\,k(x,x') < \varepsilon/\,2$ for all $x' \in G$. Thus, we obtain

$$\begin{aligned}&|M_{x,y}(x',G') - M_{x,y}(\hat{0},\{\hat{0}\})|\\ &\quad\le \left|\tfrac{k(y,x')}{k(x,x')}\,\delta_{x'}(G') + \tfrac{k(x,x')-k(y,x')}{k(x,x')}\,L_{x,y}(x',G') - \delta_{\hat{0}}(\{\hat{0}\})\right| \le 2\,\left|\tfrac{k(x,x')-k(y,x')}{k(x,x')}\right| < \varepsilon,\end{aligned}$$

for every $x' \in G'$ and for every measurable subset G' of G, as desired.
Finally, (iii) suppose that $k(x,\hat{0}) > k(y,\hat{0})$. By (R2) we can find some neighborhood G of $\hat{0}$ such that $k(x,x') - k(y,x') \ge \alpha := (k(x,\hat{0}) - k(y,\hat{0}))\,/2$ and $\left|\tfrac{k(y,x')}{k(x,x')} - \tfrac{k(y,\hat{0})}{k(x,\hat{0})}\right| < \varepsilon$ for all $x' \in G$. Then, by (3.7) we obtain

$$Q^{(x,y)}(G') = \int L_{x,y}(x',G')\,Q_{(x,y)}(dx') \ge \frac{\alpha}{1-\gamma_{x,y}} \int L_{x,y}(x',G')\,dx' \tag{3.8}$$

for every $x' \in G'$ and for every measurable subset G' of G. Since $Q^{(x,y)}(G') = 0$ in (3.8) implies that $L_{x,y}(x',G') = 0$ for a.e. $x' \in \mathcal{X}$, we derive

$$\begin{aligned}&|M_{x,y}(x',G') - M_{x,y}(\hat{0},\{\hat{0}\})|\\ &\quad\le \left|\frac{k(y,x')}{k(x,x')}\,\delta_{x'}(G') + \frac{k(x,x')-k(y,x')}{k(x,x')}\,L_{x,y}(x',G') - \frac{k(y,\hat{0})}{k(x,\hat{0})}\,\delta_{\hat{0}}(\{\hat{0}\})\right|\\ &\quad\le \left|\frac{k(y,x')}{k(x,x')} - \frac{k(y,\hat{0})}{k(x,\hat{0})}\right| < \varepsilon \quad \text{for a.e. } x' \in G',\end{aligned}$$

as desired. □

Now we assume that the Markov kernel K is absolutely continuous and stochastically monotone, and choose $\mathcal{W} := \{(x,y) : x \le y\}$. Since $K(x,\cdot) \preceq K(y,\cdot)$ implies $Q_{(x,y)}(\cdot) \preceq Q^{(x,y)}(\cdot)$, by Nachbin–Strassen theorem there exists an upward kernel $L_{x,y}$ satisfying (3.7). If we can find further that $L_{x,y}(x',A)$ is jointly measurable for each $A \in \mathcal{B}$, then the resulting γ-coupling $M_{x,y}(x',\cdot)$ is monotone and also satisfies (R3) by Proposition 3.2.

3.3 Quasi-monotone case

Given a measurable subset $\mathcal{W}$ of $\mathcal{X} \times \mathcal{X}$, we will say that a quasi-monotone case obtains, if (a) there exists a particular element in $\mathcal{X}$, say $\hat{1}$, satisfying $(x,\hat{1}) \in \mathcal{W}$ for all $x \in \mathcal{X}$, (b) there exists a particular element in $\mathcal{X}$, say $\hat{0}$, satisfying $k(x,\hat{0}) \ge k(y,\hat{0})$ for every $(x,y) \in \mathcal{W}$, and

(c) there exists a coupling $\{M_{x,y} : (x,y) \in \mathcal{W}\}$ for K in the sense of Section 3.2. Then, in this subsection we reexamine the proof of Theorem 1.2, and attempt to extend the applicability of Algorithm 1.1 for the quasi-monotone case. At the first glance of the derivation of the least upper bound c^* in (3.3), the stochastic monotonicity plays a role only in the form of Corollary 2.3. Should Corollary 2.3 be granted, one can show that $k^t(z,\hat{0}) \geq k^t(\hat{1},\hat{0})$ for all $z \in \mathcal{X}$ even for the quasi-monotone case, and therefore derive (3.3). The next lemma will show that we can generalize Corollary 2.3 for the quasi-monotone case.

Lemma 3.3. *When the quasi-monotone case obtains, we have $k^t(x,\hat{0}) \geq k^t(y,\hat{0})$ for every $(x,y) \in \mathcal{W}$.*

Proof. We can prove it by induction for $t \geq 2$. Recall the discussion in Section 2.5. Here we can similarly verify (2.9) with the bivariate Markov kernel (3.5) on $\mathcal{W}$. Since we have $k^{t-1}(x',\hat{0}) \geq k^{t-1}(y',\hat{0})$ for every $(x',y') \in \mathcal{W}$ by the induction hypothesis, we derive for each $(x,y) \in \mathcal{W}$,

$$\begin{aligned} k^t(y,\hat{0}) &= \int k^{t-1}(y',\hat{0})\, k(y,y')\, dy' = \iint k^{t-1}(y',\hat{0})\, \hat{K}((x,y), dx' \times dy') \\ &\geq \iint k^{t-1}(x',\hat{0})\, \hat{K}((x,y), dx' \times dy') = \int k^{t-1}(x',\hat{0})\, k(x,x')\, dx' = k^t(x,\hat{0}). \end{aligned}$$

□

Now it is perfectly legitimate to run Algorithm 1.1 for the quasi-monotone case, and the success of the rejection sampling method solely depends on the mechanism to generate the *correct* acceptance probability (3.4). One can construct the t-fold coupling ${M^t}_{x,y}$ recursively via (2.10) using the bivariate Markov kernel (3.5) on $\mathcal{W}$. Then, as we have demonstrated in Section 3.1, the second phase of Algorithm 1.1 yields the acceptance probability ${M^t}_{z,\hat{1}}(\hat{0},\{\hat{0}\})$.

Proposition 3.4. *Suppose that*

(R4) *the coupling $M_{x,y}(x',\cdot)$ satisfies (2.8) whenever $k(x,\hat{0}) > 0$ and $(x,y) \in \mathcal{W}$.*

Then it satisfies (2.12) whenever $k^t(x,\hat{0}) > 0$ and $(x,y) \in \mathcal{W}$.

Proof. It is exactly that of Proposition 2.6 except that we now have a rather general set $\mathcal{W}$ from our quasi-monotone case. □

Thus, we conclude that if (R1), (R2), and (R4) hold, then Algorithm 1.1 works correctly even when the quasi-monotone case is considered, and that the monotone case simply becomes a special case. Recall in Section 2.4 that the regularity conditions are sufficient for (R4). Also note that if the coupling in the quasi-monotone case is a γ-coupling, then (R4) holds automatically.

Acknowledgments. We thank James A. Fill for discussions about the content of this paper on numerous occasions, and also thank the anonymous referees for providing helpful comments and suggestions.

References

[1] Fill. J. A. (1998). An interruptible algorithm for perfect sampling via Markov chains. *Ann. Appl. Probab.* **8** 131–162.

[2] Fill, J. A. and Machida, M. (2001). Stochastic monotonicity and realizable monotonicity. *Ann. Probab.* **29** 938–978.

[3] Fill, J. A., Machida, M., Murdoch, D. J. and Rosenthal, J. S. (2000). Extension of Fill's perfect rejection sampling algorithm to general chains. *Random Structures and Algorithms* **17** 290–316.

[4] Dudley, R. M. (1989). *Real Analysis and Probability.* Wadsworth and Brooks/Cole, Pacific Grove, California.

[5] Folland, G. B. (1984). *Real Analysis.* John Wiley & Sons, New York.

[6] Kamae, T., Krengel, U., and O'Brien, G. L. (1977). Stochastic inequalities on partially ordered state spaces. *Ann. Probab.* **5** 899–912.

[7] Lindvall, T. (1992). *Lectures on the Coupling Method.* John Wiley & Sons, New York.

[8] Møller, J. and Schladitz, K. (1999). Extensions of Fill's algorithm for perfect simulation. *Journal of the Royal Statistical Society, Series B* **61** 955–969.

[9] Murdoch, D. J. and Green, P. J. (1998). Exact sampling from a continuous state space. *Scandinavian Journal of Statistics* **25** 483–502.

[10] Nachbin, L. (1965). *Topology and Order.* Van Nostrand, New York.

[11] Ross, S. (1994). *A First Course in Probability,* 6th ed. Macmillan, New York.

[12] Thönnes, E. (1999). Perfect simulation of some point processes for the impatient user. *Advances in Applied Probability* **31** 69–87.