Electronic Journal of Statistics
Vol. 10 (2016) 308–329
ISSN: 1935-7524
DOI: 10.1214/16-EJS1106

A comparison theorem for data augmentation algorithms with applications

Hee Min Choi

Department of Statistics
University of California, Davis
e-mail: hmchoi@ucdavis.edu

and

James P. Hobert

Department of Statistics
University of Florida
e-mail: jhobert@stat.ufl.edu

Abstract: The data augmentation (DA) algorithm is considered a useful Markov chain Monte Carlo algorithm that sometimes suffers from slow convergence. It is often possible to convert a DA algorithm into a sandwich algorithm that is computationally equivalent to the DA algorithm, but converges much faster. Theoretically, the reversible Markov chain that drives the sandwich algorithm is at least as good as the corresponding DA chain in terms of performance in the central limit theorem and in the operator norm sense. In this paper, we use the sandwich machinery to compare two DA algorithms. In particular, we provide conditions under which one DA chain can be represented as a sandwich version of the other. Our results are used to extend Hobert and Marchev's (2008) results on the Haar PX-DA algorithm and to improve the collapsing theorem of Liu et al. (1994) and Liu (1994). We also illustrate our results using Brownlee's (1965) stack loss data.

MSC 2010 subject classifications: Primary 60J27; secondary 62F15.
Keywords and phrases: Data augmentation algorithm, sandwich algorithm, central limit theorem, convergence rate, operator norm.

Received July 2015.

Contents

1. Introduction

Suppose f_X is an intractable density on X that we would like to explore. Consider a data augmentation (DA) algorithm (Tanner and Wong, 1987) based on the joint density $f(x,y)$ on $\mathsf{X}\times\mathsf{Y}$, which must satisfy $\int_{\mathsf{Y}} f(x,y)\,dy = f_X(x)$. The Markov chain, $\Phi = \{\Phi_m\}_{m=0}^{\infty}$, underlying the DA algorithm has the Markov transition density (Mtd) given by

$$k(x' \,|\, x) = \int_{\mathsf{Y}} f_{X|Y}(x' \,|\, y)\, f_{Y|X}(y \,|\, x)\, dy\ .$$

In other words, $k(\cdot \,|\, x)$ is the density of Φ_{m+1}, given that $\Phi_m = x$. It is well-known and easy to see that the Markov chain driven by the DA algorithm is reversible with respect to f_X, and this of course implies that f_X is an invariant density. We assume throughout this section that all Markov chains on X are Harris ergodic (see Section 2 for definition). The DA chain can be simulated by drawing alternately from the two conditional densities defined by $f(x,y)$. If the current state is $\Phi_m = x$, then Φ_{m+1} is simulated in two steps: draw $Y \sim f_{Y|X}(\cdot \,|\, x)$, call the result y, and then draw $\Phi_{m+1} \sim f_{X|Y}(\cdot \,|\, y)$.

The DA algorithm is considered a useful algorithm that sometimes suffers from slow convergence. It is often possible to convert a DA algorithm into a sandwich algorithm that is computationally equivalent to the DA algorithm, but converges much faster (see Khare and Hobert (2011), and the references therein). Let f_Y denote the y-marginal density of $f(x,y)$. The Mtd of the sandwich algorithm is given by

$$k_Q(x' \,|\, x) = \int_{\mathsf{Y}}\int_{\mathsf{Y}} f_{X|Y}(x' \,|\, y')\, Q(y, dy')\, f_{Y|X}(y \,|\, x)\, dy\ ,$$

where $Q(y, dy')$ is a Markov transition function (Mtf) on Y that is reversible with respect to f_Y. It is easy to see that $k_Q(x \,|\, x')\, f_X(x')$ is symmetric in (x, x'), so the Markov chain, $\tilde{\Phi} = \{\tilde{\Phi}_m\}_{m=0}^{\infty}$, underlying the sandwich algorithm is reversible with respect to f_X. If the current state of the sandwich chain is $\tilde{\Phi}_m = x$, then $\tilde{\Phi}_{m+1}$ can be simulated as follows. Draw $Y \sim f_{Y|X}(\cdot \,|\, x)$, call the observed value y, then draw $Y' \sim Q(y, \cdot)$, call the result y', and finally draw $\tilde{\Phi}_{m+1} \sim f_{X|Y}(\cdot \,|\, y')$. The first and third steps are exactly the two steps used to simulate the DA algorithm, and the name "sandwich algorithm", which was coined by Yu and Meng (2011), is based on the fact that the extra draw from $Q(y, \cdot)$ is sandwiched between the draws from the two conditional densities.

It is known that the sandwich chain always converges at least as fast as the DA chain in the operator norm sense. Indeed, Hobert and Román (2011) show that Yu and Meng's (2011) Theorem 1 can be used to establish that

$$\|K_Q\| \leq \|Q\|\|K\| \,,$$

where K, K_Q and Q denote the usual Markov operators defined by k, k_Q and $Q(y, dy')$, respectively, and $\|\cdot\|$ denotes the operator norm. (The operator norm will be formally defined in Section 2, but for now it suffices to note that the norm is between 0 and 1, and smaller is better.) Moreover, it follows from Hobert and Marchev's (2008) Corollary 1 that k_Q is at least as good as k in the efficiency ordering of Mira and Geyer (1999), which concerns performance in the central limit theorem (CLT).

While the sandwich machinery was designed to improve a given DA algorithm, it can also be used to compare two DA chains. In particular, suppose that, in addition to the DA algorithm based on $f(x, y)$, we have a second DA algorithm based on another joint density $\tilde{f}(x, y)$ on $\mathsf{X} \times \mathsf{Y}$ such that its x-marginal density is f_X. Suppose also that we have

$$\begin{aligned} \tilde{k}(x' \,|\, x) &= \int_{\mathsf{Y}} \tilde{f}_{X|Y}(x' \,|\, y)\, \tilde{f}_{Y|X}(y \,|\, x)\, dy \\ &= \int_{\mathsf{Y}} \int_{\mathsf{Y}} f_{X|Y}(x' \,|\, y')\, Q(y, dy')\, f_{Y|X}(y \,|\, x)\, dy \,, \end{aligned} \tag{1.1}$$

where $\tilde{f}_{X|Y}$ and $\tilde{f}_{Y|X}$ are conditional densities associated with $\tilde{f}(x, y)$ and Q is a Mtf on Y that is reversible with respect to f_Y. Then the results described above imply that $\tilde{k}$ is at least as good as k in the efficiency ordering and in the operator norm sense. The main result of this paper provides conditions under which $\tilde{k}$ admits this sandwich representation (1.1). We now provide an overview of our main result in the special case where X and Y are Euclidean spaces, and $f(x, y)$ and $\tilde{f}(x, y)$ are densities with respect to Lebesgue measure.

Let $\tilde{f}_Y$ denote the y-marginal density of $\tilde{f}(x, y)$. Suppose that there exits a Mtf R on Y satisfying

$$\tilde{f}_{X|Y}(x \,|\, y) = \int_{\mathsf{Y}} f_{X|Y}(x \,|\, y')\, R(y, dy')$$

and

$$\int_{y \in \mathsf{Y}} R(y, dy')\, \tilde{f}_Y(y)\, dy = f_Y(y')\, dy' \,.$$

Then (1.1) holds with Q equal to the Mtf corresponding to the DA algorithm for f_Y based on the joint distribution on $\mathsf{Y} \times \mathsf{Y}$ given by $R(y, dy')\, \tilde{f}_Y(y)\, dy$.

We now illustrate the use of our results with several applications. Our main application involves generalizing the results of Hobert and Marchev (2008) (hereafter, H&M), who themselves generalized results of Liu and Wu (1999) (hereafter, L&W). L&W developed the PX-DA algorithm. The basic idea is to use

$f(x,y)$ to create an entire family of joint densities on $\mathsf{X}\times\mathsf{Y}$ such that the x-marginal density of each member is f_X. This allows for the construction of a class of viable DA algorithms. To be specific, consider a class of functions $t_g:\mathsf{Y}\to\mathsf{Y}$ for $g\in G$ such that, for each fixed g, $t_g(y)$ is one-to-one and differentiable in y. We are assuming here that G is a group with identity element e. Assume further that (a) $t_e(y)=y$ for all $y\in\mathsf{Y}$ and (b) $t_{g_1g_2}(y)=t_{g_1}(t_{g_2}(y))$ for $g_1,g_2\in G$ and all $y\in\mathsf{Y}$. Suppose that $\nu: G\times\mathsf{X}\to[0,\infty)$ is a conditional probability density with respect to (unimodular) Haar measure ϱ (see Section 3 for definition). Now, define a probability density $\tilde{f}^{(\nu)}(x,y,g)=f(x,t_g(y))\,|J_g(y)|\,\nu(g\,|\,x)$ on $\mathsf{X}\times\mathsf{Y}\times G$, where $J_g(z)$ is the Jacobian of the transformation $z=t_g^{-1}(y)$. Let

$$\tilde{f}^{(\nu)}(x,y)=\int_G \tilde{f}^{(\nu)}(x,y,g)\,\varrho(dg)\ .$$

Clearly, the x-marginal of $\tilde{f}^{(\nu)}(x,y)$ is the target, f_X. Thus, each conditional density $\nu(g\,|\,x)$ leads to a new DA algorithm. L&W also propose the Haar PX-DA algorithm, which is a popular sandwich algorithm where $Q(y,dy')$ corresponds to the move $y\to y'=t_g(y)$ with g (on G) drawn from the density (with respect to ϱ) proportional to $f_Y(t_g(y))\,|J_g(y)|$.

L&W establish that the Haar PX-DA algorithm is at least as good in the operator norm sense as every PX-DA algorithm in the special case where X,Y and G are Euclidean spaces and the group G is unimodular. H&M provide extensions and generalizations of L&W's results in the special case where $\nu(\cdot\,|\,x)$ does not depend on x. In particular, H&M show that L&W's results hold on more general spaces, and that Haar PX-DA is also at least as good as PX-DA in the efficiency ordering. Moreover, H&M are able to remove a key regularity condition that is required by L&W. In our main application, we show that all of H&M's results continue to hold in the more general case where $\nu(\cdot\,|\,x)$ does depend on x.

We also apply our results to improve the collapsing theorem (Liu et al., 1994; Liu, 1994). To be specific, suppose there exists a joint density $f(x,y,z)$ on $\mathsf{X}\times\mathsf{Y}\times\mathsf{Z}$ such that $\int_{\mathsf{Z}} f(x,y,z)\,dz=f(x,y)$. Liu et al. (1994) refer to the DA algorithm which iterates between drawing $f_{Y,Z|X}$ and drawing $f_{X|Y,Z}$ as "grouping" and the DA algorithm based on $f(x,y)$ as "collapsing". The collapsing theorem implies that collapsing DA converges at least as fast as grouping DA in the operator norm sense. We show that the collapsing DA chain is also at least as good as the grouping DA chain in the efficiency ordering.

The remainder of this paper is organized as follows. Section 2 contains some results from general state space Markov chain theory, our main result, and a toy example. In Section 3, we apply the main result to extend H&M's results on the Haar PX-DA. We also illustrate our results using a PX-DA algorithm and Choi and Hobert's (2013) Haar PX-DA algorithm for Bayesian linear regression with Laplace errors. Finally, our main result is used to improve the collapsing theorem in Section 4.

2. A comparison theorem

2.1. Markov chain background

Let $P(x, dx')$ be a Mtf on a topological space X equipped with its Borel σ-algebra $\mathcal{B}_{\mathsf{X}}$. Suppose $P(x, dx')$ is reversible with respect to a probability measure π. Denote the Markov chain defined by $P(x, dx')$ as $\Phi = \{\Phi_m\}_{m=0}^{\infty}$. Assume that Φ is Harris ergodic (i.e., irreducible, aperiodic, and positive Harris recurrent). As usual, let $L^2(\pi)$ be the vector space of real-valued, measurable functions on X that are square-integrable with respect to π, and let $L_0^2(\pi)$ be the subspace of mean-zero functions. This is a Hilbert space in which inner product of $g, h \in L_0^2(\pi)$ is defined as

$$\langle g, h \rangle = \int_{\mathsf{X}} g(x)\, h(x)\, \pi(dx) \;,$$

and the corresponding norm is, of course, given by $\|g\| = \langle g, g \rangle^{1/2}$. The Mtf $P(x, dx')$ defines an operator P on $L_0^2(\pi)$ that maps $g \in L_0^2(\pi)$ to

$$(Pg)(x) = \int_{\mathsf{X}} g(x')\, P(x, dx') = \mathrm{E}\left[g(\Phi_{m+1}) \,\middle|\, \Phi_m = x \right] \;.$$

It is easy to see, using reversibility, that P is self-adjoint; that is, for all $g, h \in L_0^2(\pi)$, $\langle Pg, h \rangle = \langle g, Ph \rangle$. The operator norm of P is defined as

$$\|P\| = \sup_{\{g \in L_0^2(\pi) : \|g\| = 1\}} \|Pg\| \;.$$

A simple application of Jensen's inequality shows that $\|P\| \in [0, 1]$. In fact, $\|P\|$ provides a great deal of information about the convergence behavior of the corresponding Markov chain Φ. For instance, Φ is geometrically ergodic if and only if $\|P\| < 1$ (Roberts and Rosenthal, 1997). Moreover, results in Liu et al. (1995) show that the smaller the norm, the faster the chain converges.

Assume that Markov chain Monte Carlo will be used to estimate the finite, intractable expectation $E_\pi g = \int_{\mathsf{X}} g(x)\, \pi(dx)$. Assume further that there exists a CLT for the ergodic average $\bar{g}_m = \frac{1}{m} \sum_{i=0}^{m-1} g(\Phi_i)$; that is, there exists $\sigma_g^2 \in (0, \infty)$ such that, as $m \to \infty$, $\sqrt{m}(\bar{g}_m - E_\pi g) \to N(0, \sigma_g^2)$ in distribution. (If there is no CLT, then we simply write $\sigma_g^2 = \infty$.) Suppose we have two Harris ergodic Mtfs P and Q that have π as an invariant probability measure. Denote σ_g^2 for the two Mtfs by $\sigma_g^2(P)$ and $\sigma_g^2(Q)$. We say P is at least as good as Q in the efficiency ordering, written $P \succeq_E Q$, if $\sigma_g^2(P) \leq \sigma_g^2(Q)$ for every $g \in L^2(\pi)$ (Mira and Geyer, 1999).

2.2. The main result

Let X and Y be separable metric spaces equipped with their Borel σ-algebras. We will refer to such a space Y as a *sub-Cauchy space* if there exists a complete separable metric space $\bar{\mathsf{Y}}$ such that Y is a Borel subset of $\bar{\mathsf{Y}}$. We assume that Y

is sub-Cauchy. This is a weak assumption. For example, with Euclidean metric, if $\mathsf{Y} = \mathbb{R}^d$, then $\bar{\mathsf{Y}} = \mathbb{R}^d$, and if $\mathsf{Y} = (0,\infty)^d$, then $\bar{\mathsf{Y}} = [0,\infty)^d$. Assume further that μ_X and μ_Y are σ-finite measures on X and Y, and that $f(x,y)$ and $\tilde{f}(x,y)$ are two different probability densities on $\mathsf{X}\times\mathsf{Y}$ with respect to $\mu_X \times \mu_Y$ such that

$$\int_{\mathsf{Y}} f(x,y)\,\mu_Y(dy) = \int_{\mathsf{Y}} \tilde{f}(x,y)\,\mu_Y(dy) = f_X(x)\ .$$

In this context, the Mtd of the DA chain based on the joint density $f(x,y)$ is

$$k(x\,|\,x') = \int_{\mathsf{Y}} f_{X|Y}(x\,|\,y)\, f_{Y|X}(y\,|\,x')\,\mu_Y(dy)\ ,$$

where $f_{X|Y}$ and $f_{Y|X}$ are the conditional densities associated with $f(x,y)$. Analogously, let $\tilde{k}$ be the Mtd of the DA chain for f_X based on the joint density $\tilde{f}(x,y)$. As usual, let f_Y denote the y-marginal density of $f(x,y)$, and let $\tilde{f}_Y$ and $\tilde{f}_{X|Y}$ be the marginal and conditional densities defined by $\tilde{f}(x,y)$. The following result allows us to compare the two DA chains.

Theorem 2.1. *Let K and $\tilde{K}$ denote the operators defined by k and $\tilde{k}$. Assume that the Markov chains driven by k and $\tilde{k}$ are Harris ergodic. If there exists a Mtf R on Y satisfying*

$$\tilde{f}_{X|Y}(x\,|\,y) = \int_{\mathsf{Y}} f_{X|Y}(x\,|\,y')\,R(y,dy')\ ,$$

and

$$\int_{y\in\mathsf{Y}} R(y,dy')\,\tilde{f}_Y(y)\,\mu_Y(dy) = f_Y(y')\,\mu_Y(dy')\ ,$$

then $\|\tilde{K}\| \le \gamma^2\|K\|$ and $\tilde{k} \succeq_E k$, where γ is the maximal correlation of the pair (Y,Y') whose joint distribution is $R(y,dy')\,\tilde{f}_Y(y)\,\mu_Y(dy)$.

Proof. Our assumptions about the space Y imply that every probability measure on Y is tight (see Parthasarathy, 1967, Theorem 3.2). It then follows from Theorem 6 of Faden (1985) and Theorem 2.4.1 of Ramachandran (1979) that there exists a Mtf R^* on Y such that, for all $y, y'' \in \mathsf{Y}$,

$$R^*(y,dy'')\,f_Y(y)\,\mu_Y(dy) = R(y'',dy)\,\tilde{f}_Y(y'')\,\mu_Y(dy'')\ . \tag{2.1}$$

We now show that $\tilde{k}$ can be written as

$$\begin{aligned}\tilde{k}(x\,|\,x') &= \int_{\mathsf{Y}} \tilde{f}_{X|Y}(x\,|\,y)\,\tilde{f}_{Y|X}(y\,|\,x')\,\mu_Y(dy)\\ &= \int_{\mathsf{Y}}\int_{\mathsf{Y}} f_{X|Y}(x\,|\,y')\,Q(y,dy')\,f_{Y|X}(y\,|\,x')\,\mu_Y(dy)\ ,\end{aligned}$$

where

$$Q(y,dy') = \int_{y''\in\mathsf{Y}} R(y'',dy')\,R^*(y,dy'')\ . \tag{2.2}$$

Indeed,

$$\begin{aligned}
\tilde{k}(x \,|\, x') &= \int_{\mathsf{Y}} \tilde{f}_{X|Y}(x \,|\, y'')\, \tilde{f}_{Y|X}(y'' \,|\, x')\, \mu_Y(dy'') \\
&= \int_{\mathsf{Y}} \left[\int_{\mathsf{Y}} f_{X|Y}(x \,|\, y')\, R(y'', dy') \right] \\
&\qquad \times \left[\frac{\tilde{f}_Y(y'')}{f_X(x')} \int_{\mathsf{Y}} f_{X|Y}(x' \,|\, y)\, R(y'', dy) \right] \mu_Y(dy'') \\
&= \int_{\mathsf{Y}} \left[\int_{\mathsf{Y}} \frac{\left[\int_{\mathsf{Y}} f_{X|Y}(x \,|\, y')\, R(y'', dy') \right] f_{Y|X}(y \,|\, x')}{f_Y(y)} R(y'', dy) \right] \\
&\qquad \times \tilde{f}_Y(y'')\, \mu_Y(dy'') \\
&= \int_{\mathsf{Y}} \int_{\mathsf{Y}} f_{X|Y}(x \,|\, y') \left[\int_{y'' \in \mathsf{Y}} R(y'', dy')\, R^*(y, dy'') \right] f_{Y|X}(y \,|\, x')\, \mu_Y(dy) \\
&= \int_{\mathsf{Y}} \int_{\mathsf{Y}} f_{X|Y}(x \,|\, y')\, Q(y, dy')\, f_{Y|X}(y \,|\, x')\, \mu_Y(dy) \,,
\end{aligned}$$

where the penultimate equality follows from (2.1). Here, the Mtf $Q(y, dy')$ is reversible with respect to f_Y since (2.2) indicates that the Markov chain defined by $Q(y, dy')$ is a DA for f_Y based on the joint distribution (2.1). An application of Hobert and Marchev's (2008) Corollary 1 implies $\tilde{k} \succeq_E k$. Moreover, it follows from Hobert and Román (2011) that $\|\tilde{K}\| \leq \|Q\|\|K\|$, where Q is the operator on $L_0^2(f_Y)$ corresponding to $Q(y, dy')$. By Liu et al.'s (1994) Theorem 3.2, $\|Q\| = \gamma^2$. The proof is complete. □

Remark 2.1. We note that, under the conditions in Theorem 2.1, $\tilde{k}$ is actually the Mtd of a GIS algorithm based on $f, \tilde{f}$ and R^* (see Yu and Meng, 2011; Hobert and Román, 2011). Indeed, by (2.1), we have

$$\begin{aligned}
\tilde{k}(x \,|\, x') &= \int_{\mathsf{Y}} \tilde{f}_{X|Y}(x \,|\, y'')\, \tilde{f}_{Y|X}(y'' \,|\, x')\, \mu_Y(dy'') \\
&= \int_{\mathsf{Y}} \int_{\mathsf{Y}} \tilde{f}_{X|Y}(x \,|\, y'') \frac{f_{X|Y}(x' \,|\, y)}{f_X(x')} R(y'', dy)\, \tilde{f}_Y(y'')\, \mu_Y(dy'') \\
&= \int_{\mathsf{Y}} \int_{\mathsf{Y}} \tilde{f}_{X|Y}(x \,|\, y'') \frac{f_{X|Y}(x' \,|\, y)}{f_X(x')} R^*(y, dy'')\, f_Y(y)\, \mu_Y(dy) \\
&= \int_{\mathsf{Y}} \int_{\mathsf{Y}} \tilde{f}_{X|Y}(x \,|\, y'')\, R^*(y, dy'') f_{Y|X}(y \,|\, x')\, \mu_Y(dy) \,.
\end{aligned}$$

Another application of (2.1) reveals that

$$\int_{y \in \mathsf{Y}} R^*(y, dy')\, f_Y(y)\, \mu_Y(dy) = \tilde{f}_Y(y')\, \mu_Y(dy') \int_{\mathsf{Y}} R(y', dy) = \tilde{f}_Y(y')\, \mu_Y(dy') \,.$$

This GIS representation suggests that, if we inappropriately design a GIS algorithm, then we could end up with one of the original DA algorithms.

Remark 2.2. When comparing DA algorithms, computing time and simulation efforts should be taken into account in addition to the efficiency and speed of convergence, but we will not get into that here.

2.3. A toy example

Consider the following well-known toy example involving a simple two-level normal hierarchical linear model (see e.g., Liu and Wu, 1999; Yu and Meng, 2011)

$$\begin{aligned} Y \,|\, (\theta, Z) &\sim N(\theta + Z, 1)\ , \\ Z \,|\, \theta &\sim N(0, D)\ . \end{aligned} \tag{2.3}$$

Here, θ is the parameter, Y is the observed data, Z is the latent variable, and D is known positive constant. We assume that $D \neq 1$. With a flat improper prior on θ, the posterior is $\theta \,|\, Y \sim N(Y, 1 + D)$, which is our target density. Treating Z as the latent variable, the standard DA algorithm is simulated by drawing alternately from the following two conditional distributions:

$$\begin{aligned} Z \,|\, (\theta, Y) &\sim N\left(\frac{D(Y-\theta)}{1+D}, \frac{D}{1+D}\right)\ , \\ \theta \,|\, (Z, Y) &\sim N(Y - Z, 1)\ . \end{aligned}$$

On the other hand, if we let $\tilde{Z} = Z + \theta$, and treat $\tilde{Z}$ as the latent data, then the model can be rewritten as

$$\begin{aligned} Y \,|\, (\theta, \tilde{Z}) &\sim N(\tilde{Z}, 1)\ , \\ \tilde{Z} \,|\, \theta &\sim N(\theta, D)\ . \end{aligned}$$

This is called the centered parametrization (CP), whereas model (2.3) is called the non-centered parametrization (NCP). If we put the same flat prior on θ, then this model leads to a different DA algorithm, which iterates between drawing $\tilde{Z} \,|\, (\theta, Y)$ and drawing $\theta \,|\, (\tilde{Z}, Y)$:

$$\begin{aligned} \tilde{Z} \,|\, (\theta, Y) &\sim N\left(\frac{\theta + DY}{1+D}, \frac{D}{1+D}\right)\ , \\ \theta \,|\, (\tilde{Z}, Y) &\sim N(\tilde{Z}, D)\ . \end{aligned}$$

Though both DAs have the same target distribution, they have completely different convergence behavior. Let k and $\tilde{k}$ denote Mtds of NCP and CP DA chains, and let K and $\tilde{K}$ denote operators associated with k and $\tilde{k}$. It is known that $\|K\| = \frac{D}{1+D}$ and $\|\tilde{K}\| = \frac{1}{1+D}$. Therefore, when $D > 1$, the CP DA algorithm dominates NCP DA algorithm in the operator norm sense. On the other hand, when $D < 1$, the operator norm ordering is reversed. Using Theorem 2.1, we can show similar ordering results hold in terms of efficiency. Here is the result.

Proposition 2.1. *When $D > 1$, $\tilde{k} \succeq_E k$. When $D < 1$, $k \succeq_E \tilde{k}$.*

Proof. Let $f(\theta, z \,|\, y)$ denote the density of $(\theta, Z) \,|\, Y$ in the above NCP model, and let $f_{\theta|Z,Y}$ and $f_{Z|Y}$ be the conditional and marginal densities defined by $f(\theta, z \,|\, y)$. Similarly, denote the density of $(\theta, \tilde{Z}) \,|\, Y$ in the CP model as $\tilde{f}(\theta, \tilde{z} \,|\, y)$, and let $\tilde{f}_{\theta|\tilde{Z},Y}$ and $\tilde{f}_{\tilde{Z}|Y}$ be the associated conditional and marginal densities. It is easy to see that $f_{Z|Y}(\cdot \,|\, y) \sim N(0, D)$ and $\tilde{f}_{\tilde{Z}|Y}(\cdot \,|\, y) \sim N(y, 1)$. We begin with the case where $D > 1$. It suffices to establish the two conditions of Theorem 2.1. Let $r(w \,|\, z, y)$ denote the $N(y - z, D - 1)$ density evaluated at w. It is easy to show that

$$\begin{aligned}\int_{\mathbb{R}} f_{\theta|Z,Y}(\theta \,|\, z', y)\, r(z' \,|\, z, y)\, dz' &= (2\pi D)^{-\frac{1}{2}} \exp\left\{-\frac{1}{D}(\theta - z)^2\right\} \\ &= \tilde{f}_{\theta|\tilde{Z},Y}(\theta \,|\, z, y)\end{aligned}$$

and

$$\int_{\mathbb{R}} r(z' \,|\, z, y)\, \tilde{f}_{\tilde{Z}|Y}(z \,|\, y)\, dz = (2\pi D)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2D} z'^2\right\} = f_{Z|Y}(z' \,|\, y)\ .$$

An application of Theorem 2.1 yields the result.

We now prove the case where $D < 1$. Similarly, we establish the two conditions of Theorem 2.1. Let $\tilde{r}(w \,|\, z, y)$ denote the $N(y - z, 1 - D)$ density evaluated at w. Then, we have

$$\begin{aligned}\int_{\mathbb{R}} \tilde{f}_{\theta|\tilde{Z},Y}(\theta \,|\, z', y)\, \tilde{r}(z' \,|\, z, y)\, dz' &= (2\pi)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\theta - y + z)^2\right\} \\ &= f_{\theta|Z,Y}(\theta \,|\, z, y)\end{aligned}$$

and

$$\int_{\mathbb{R}} \tilde{r}(z' \,|\, z, y)\, f_{Z|Y}(z \,|\, y)\, dz = (2\pi)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(z' - y)^2\right\} = \tilde{f}_{\tilde{Z}|Y}(z' \,|\, y)\ .$$

Another application of Theorem 2.1 completes the proof. □

It is interesting to compare the exact value of $\|\tilde{K}\|$ with the upper bound $\gamma^2\|K\|$ from Theorem 2.1. Consider the case where $D > 1$. We know from the proof of Proposition 2.1 that γ is equal to the maximal correlation of a random pair (Z, Z') with joint density $r(z' \,|\, z, y)\, \tilde{f}_{Z|Y}(z \,|\, y)$. This joint density is bivariate normal, and it follows from Gebelein (1941) and Lancaster (1957) that

$$\gamma = \sqrt{\frac{\mathrm{Var}\left(\mathrm{E}(Z' \,|\, Z, Y) \big| Y\right)}{\mathrm{Var}(Z' \,|\, Y)}} = \sqrt{\frac{1}{1 + D - 1}} = \frac{1}{\sqrt{D}}\ .$$

Therefore,

$$\gamma^2\|K\| = \frac{1}{D} \cdot \frac{D}{1 + D} = \|\tilde{K}\|\ .$$

This example shows that the bound in Theorem 2.1 is tight.

3. Extending Hobert and Marchev's results on Haar PX-DA

In this section, we use Theorem 2.1 to show that H&M's extensions of L&W's results continue to hold when ν does depend on x. We also illustrate our results using Brownlee's (1965) stack loss data.

3.1. Hobert and Marchev's group structure

Let $\mathsf{X}, \mathsf{Y}, \mu_X$ and μ_Y be as in Section 2.2, and let G be a group with identity element e. Allow the group G to act topologically on the left of Y; that is, there is a continuous function $F : G \times \mathsf{Y} \to \mathsf{Y}$ such that $F(e, y) = y$ for all $y \in \mathsf{Y}$ and $F(g_1 g_2, y) = F(g_1, F(g_2, y))$ for all $g_1, g_2 \in G$ and all $y \in \mathsf{Y}$. As is typically done, we will denote the value of F at (g, y) by gy so, in this notation, the two conditions are written $ey = y$ and $(g_1 g_2)y = g_1(g_2 y)$.

Assume that there exists a function $j : G \times \mathsf{Y} \to (0, \infty)$ such that:

1. $j(g^{-1}, y) = \frac{1}{j(g,y)}$ for all $g \in G$ and all $y \in \mathsf{Y}$,
2. $j(g_1 g_2, y) = j(g_1, g_2 y)\, j(g_2, y)$ for all $g_1, g_2 \in G$ and all $y \in \mathsf{Y}$, and
3. For all $g \in G$ and all integrable functions $h : \mathsf{Y} \to \mathbb{R}$,

$$\int_{\mathsf{Y}} h(gy)\, j(g, y)\, \mu_Y(dy) = \int_{\mathsf{Y}} h(y)\, \mu_Y(dy) \, .$$

As in L&W, suppose that $\mathsf{Y} \subseteq \mathbb{R}^n$, μ_Y is Lebesgue measure on Y, and for each fixed $g \in G$, $F(g, \cdot) : \mathsf{Y} \to \mathsf{Y}$ is differentiable. Then if we take $j(g, y)$ to be the Jacobian of the transformation $y \mapsto F(g, y)$, the three properties listed above can be easily verified from calculus.

3.2. Constructing general PX-DA and Haar PX-DA algorithms

As before, let $f(x, y)$ be a probability density on $\mathsf{X} \times \mathsf{Y}$ with respect to $\mu_X \times \mu_Y$ whose x-marginal density is f_X. We construct a general PX-DA algorithm for f_X. The idea is to build a joint density, that is a general version of $\tilde{f}^{(\nu)}(x, y)$ of L&W's PX-DA described in the introduction, using the group structure on G. This leads to a new DA algorithm. To be specific, define a probability density on $\mathsf{X} \times \mathsf{Y}$ as follows:

$$\tilde{f}^{(\nu)}(x, y) = \int_G f(x, gy)\, j(g, y)\, \nu(x, dg) \, , \tag{3.1}$$

where $\nu(x, \cdot)$ is a conditional probability measure on G given $x \in \mathsf{X}$. It is easy to see that the x-marginal density of $\tilde{f}^{(\nu)}(x, y)$ is f_X, and the y-marginal density is

$$m_\nu(y) := \int_{\mathsf{X}} \left[\int_G f(x, gy)\, j(g, y)\, \nu(x, dg) \right] \mu_X(dx) \, ,$$

where we assume $m_\nu(y)$ is positive, finite for all $y \in \mathsf{Y}$. (As in H&M, it is possible to handle cases where $m_\nu(y) = \infty$ on a set of Y that has μ_Y-measure zero, but we will not go into that here.) The associated conditional densities are

$$\tilde{f}^{(\nu)}_{X|Y}(x \,|\, y) = \frac{\tilde{f}^{(\nu)}(x,y)}{m_\nu(y)} = \frac{\int_G f(x,g'y)\, j(g',y)\, \nu(x,dg')}{m_\nu(y)}$$

and $\tilde{f}^{(\nu)}_{Y|X}(y \,|\, x) = \int_G f_{Y|X}(gy \,|\, x)\, j(g,y)\, \nu(x,dg)$. Our general PX-DA is a DA with Mtd given by

$$k_\nu(x \,|\, x') = \int_{\mathsf{Y}} \tilde{f}^{(\nu)}_{X|Y}(x \,|\, y)\, \tilde{f}^{(\nu)}_{Y|X}(y \,|\, x')\, \mu_Y(dy)\ .$$

We note that if $\nu(x,\cdot)$ is free of x, then we recover H&M's general PX-DA chain.

We now describe H&M's general Haar PX-DA algorithm. H&M use the group structure to construct a sandwich step, that behaves like a generalized version of the sandwich step of L&W's Haar PX-DA described in Section 1. Under the assumptions in the previous section, there exists a left-Haar measure, ϱ_l, on G, which is a nontrivial measure satisfying

$$\int_G h(\tilde{g}g)\, \varrho_l(dg) = \int_G h(g)\, \varrho_l(dg) \tag{3.2}$$

for all $\tilde{g} \in G$ and all integrable functions $h : G \to \mathbb{R}$. This measure is unique up to a multiplicative constant. Moreover, there exists a multiplier, $\triangle$, called the (right) modular function of the group, which relates the left-Haar and right-Haar measures, ϱ_l and ϱ_r, on G such that $\varrho_r(dg) = \triangle(g^{-1})\varrho_l(dg)$. (A function $\chi : G \to (0,\infty)$ is called a multiplier if χ is continuous and $\chi(g_1g_2) = \chi(g_1)\,\chi(g_2)$ for all $g_1, g_2 \in G$.) Here, the right-Haar measure satisfies the obvious analogue of (3.2). Groups for which $\triangle \equiv 1$; that is, for which right-Haar and left-Haar measures are equivalent, are called unimodular. We now state two useful formulas from H&M that will be used later. If $\tilde{g} \in G$ and $h : G \to \mathbb{R}$ is an integrable function, then

$$\int_G h(g\tilde{g}^{-1})\, \varrho_l(dg) = \triangle(\tilde{g}) \int_G h(g)\, \varrho_l(dg) \tag{3.3}$$

and

$$\int_G h(g^{-1})\, \varrho_l(dg) = \int_G h(g)\, \triangle(g^{-1})\, \varrho_l(dg)\ . \tag{3.4}$$

As before, let f_Y be the y-marginal density of $f(x,y)$, and assume without loss of generality that

$$m(y) := \int_G f_Y(gy)\, j(g,y)\, \varrho_l(dg)$$

is positive, finite for all $y \in \mathsf{Y}$. A straightforward application of (3.3) shows that, for $y \in \mathsf{Y}$,

$$m(gy) = j(g^{-1},y)\, \triangle(g^{-1})\, m(y)\ . \tag{3.5}$$

We now describe a recipe for using the group structure to build a Mtf that is reversible with respect to f_Y. Let R be an operator on $L_0^2(f_Y)$ that maps $h \in L_0^2(f_Y)$ to

$$(Rh)(y) = \int_G \frac{h(gy)\, f_Y(gy)\, j(g,y)}{m(y)} \varrho_l(dg) \; . \tag{3.6}$$

Then the corresponding Markov chain on Y evolves as follows. If the current state is y, then the distribution of the next state is that of gy where g is a random element drawn from the density (with respect to ϱ_l) $f_Y(gy)\, j(g,y)/m(y)$. We denote its Mtf by $R(y, dy')$. It is shown in H&M's Proposition 3 and 4 that the operator R on $L_0^2(f_Y)$ is self-adjoint (with respect to f_Y) and idempotent. The Mtd of H&M's general Haar PX-DA is

$$k^*(x \,|\, x') = \int_{\mathsf{Y}} \int_{\mathsf{Y}} f_{X|Y}(x \,|\, y')\, R(y, dy')\, f_{Y|X}(y \,|\, x')\, \mu_Y(dy) \; .$$

Together, H&M's Proposition 1 and Theorem 4 imply that k^* is itself a DA algorithm. Precisely, k^* is a DA algorithm based on the joint density

$$f^*(x,y) = f_Y(y) \int_{\mathsf{Y}} f_{X|Y}(x \,|\, y')\, R(y, dy') \; .$$

That is, k^* can be written as

$$\begin{aligned} k^*(x \,|\, x') &= \int_{\mathsf{Y}} \int_{\mathsf{Y}} f_{X|Y}(x \,|\, y')\, R(y, dy')\, f_{Y|X}(y \,|\, x')\, \mu_Y(dy) \\ &= \int_{\mathsf{Y}} f^*_{X|Y}(x \,|\, y)\, f^*_{Y|X}(y \,|\, x')\, \mu_Y(dy) \; , \end{aligned}$$

where $f^*_{X|Y}$ and $f^*_{Y|X}$ are conditional densities associated with $f^*(x,y)$.

3.3. Comparing General PX-DA and Haar PX-DA Algorithms

In this section, we establish that k^* is at least as good as k_ν in the efficiency ordering and in the operator norm sense. In fact, H&M's Theorem 4 implies that their general Haar PX-DA algorithm is at least as good as their general PX-DA algorithm in the efficiency ordering and operator norm sense. Since H&M's general PX-DA is a special case of our general PX-DA, our result improves upon their result. Here is our result.

Proposition 3.1. *Let $\nu(x, \cdot)$ be a conditional probability measure on G given x ($\in \mathsf{X}$). Assume that $m_\nu(y)$ and $m(y)$ are positive and finite for all $y \in \mathsf{Y}$. If the Markov chains driven by k_ν and k^* are Harris ergodic, then $k^* \succeq_E k_\nu$ and $\|K^*\| \leq \|K_\nu\|$, where K_ν and K^* are the operators on $L_0^2(f_X)$ defined by k_ν and k^*.*

Proof. Recall that k_ν is the DA Mtd based on the joint density

$$\tilde{f}^{(\nu)}(x,y) = \int_G f(x, g'y)\, j(g', y)\, \nu(x, dg')$$

and that the y-marginal density of $\tilde{f}^{(\nu)}(x,y)$ is

$$m_\nu(y) = \int_{\mathsf{X}} \left[\int_G f(x, g'y)\, j(g', y)\, \nu(x, dg') \right] \mu_X(dx) \, .$$

Let $\tilde{R}(y, dy')$ be the Mtf on Y with invariant density $m_\nu(y)$ that is constructed according to the recipe in (3.6); that is, $\tilde{R}$ is what we would have ended up with had we used $m_\nu(y)$ in place of $f_Y(y)$. If we substitute $m_\nu(y)$ for $f_Y(y)$ in the definition of $m(y)$, we have

$$\begin{aligned}
&\int_G m_\nu(gy)\, j(g,y)\, \varrho_l(dg) \\
&\quad = \int_G \left[\int_{\mathsf{X}} \int_G f(x, g'gy)\, j(g', gy)\, \nu(x, dg') \mu_X(dx) \right] j(g,y)\, \varrho_l(dg) \\
&\quad = \int_{\mathsf{X}} \int_G \left[\int_G f(x, g'gy)\, j(g'g, y)\, \varrho_l(dg) \right] \nu(x, dg')\, \mu_X(dx) \\
&\quad = \int_{\mathsf{X}} \int_G \left[\int_G f(x, gy)\, j(g,y)\, \varrho_l(dg) \right] \nu(x, dg')\, \mu_X(dx) \\
&\quad = \int_G \left[\int_{\mathsf{X}} f(x, gy)\, j(g,y)\, \mu_X(dx) \right] \varrho_l(dg) \\
&\quad = \int_G f_Y(gy)\, j(g,y)\, \varrho_l(dg) = m(y) \, .
\end{aligned}$$

Hence, the function $m(y)$ is the same whether we use f_Y or m_ν. Recall that k^* is the DA Mtd based on the joint density

$$f^*(x,y) = f_Y(y) \int_{\mathsf{Y}} f_{X|Y}(x \,|\, y')\, R(y, dy') \, .$$

Clearly, $f^*_Y(y) = \int_{\mathsf{X}} f^*(x,y)\, \mu_X(dx) = f_Y(y)$, so

$$f^*_{X|Y}(x \,|\, y) = \int_{\mathsf{Y}} f_{X|Y}(x \,|\, y')\, R(y, dy') \, .$$

We now establish the two conditions of Theorem 2.1. We will first show that

$$f^*_{X|Y}(x \,|\, y) = \int_{\mathsf{Y}} \tilde{f}^{(\nu)}_{X|Y}(x \,|\, y')\, \tilde{R}(y, dy') \, .$$

Indeed, using the definition of $\tilde{R}$ and calculation above, we have

$$\begin{aligned}
&\int_{\mathsf{Y}} \tilde{f}^{(\nu)}_{X|Y}(x \,|\, y')\, \tilde{R}(y, dy') \\
&\quad = \int_G \tilde{f}^{(\nu)}_{X|Y}(x \,|\, gy) \frac{m_\nu(gy)\, j(g,y)}{m(y)} \rho_l(dg) \\
&\quad = \int_G \int_G \frac{f(x, g'gy)\, j(g', gy)}{m_\nu(gy)} \frac{m_\nu(gy)\, j(g,y)}{m(y)} \rho_l(dg)\, \nu(x, dg')
\end{aligned}$$

$$
\begin{aligned}
&= \int_G \int_G \frac{f(x, g'gy)\, j(g'g, y)}{m(y)} \rho_l(dg)\, \nu(x, dg') \\
&= \int_G \int_G \frac{f(x, gy)\, j(g, y)}{m(y)} \rho_l(dg)\, \nu(x, dg') \\
&= \int_G f_{X|Y}(x \,|\, gy) \frac{f_Y(gy)\, j(g, y)}{m(y)} \rho_l(dg) \\
&= \int_{\mathsf{Y}} f_{X|Y}(x \,|\, y')\, R(y, dy') = f^*_{X|Y}(x \,|\, y) \; .
\end{aligned}
$$

We proceed by demonstrating that, for $y, y' \in \mathsf{Y}$,

$$
\tilde{R}(y, dy')\, f_Y(y)\, \mu_Y(dy) = R(y', dy)\, m_\nu(y')\, \mu_Y(dy') \tag{3.7}
$$

to establish

$$
\int_{y \in \mathsf{Y}} \tilde{R}(y, dy')\, f_Y(y)\, \mu_Y(dy) = m_\nu(y')\, \mu_Y(dy') \; .
$$

It suffices to show that, for bounded functions h_1, h_2 on Y,

$$
\int_{\mathsf{Y}} (\tilde{R}h_1)(y)\, h_2(y)\, f_Y(y)\, \mu_Y(dy) = \int_{\mathsf{Y}} (Rh_2)(y)\, h_1(y)\, m_\nu(y)\, \mu_Y(dy) \; .
$$

Indeed,

$$
\begin{aligned}
&\int_{\mathsf{Y}} (\tilde{R}h_1)(y)\, h_2(y)\, f_Y(y)\, \mu_Y(dy) \\
&\quad = \int_G \int_{\mathsf{Y}} \frac{h_2(y)\, h_1(gy)\, m_\nu(gy)\, j(g, y)}{m(y)} f_Y(y)\, \mu_Y(dy)\, \rho_l(dg) \\
&\quad = \int_G \int_{\mathsf{Y}} \frac{h_2(g^{-1}y)\, h_1(y)\, m_\nu(y)\, f_Y(g^{-1}y)}{m(g^{-1}y)} \mu_Y(dy)\, \rho_l(dg) \\
&\quad = \int_{\mathsf{Y}} \int_G \frac{h_2(g^{-1}y)\, h_1(y)\, m_\nu(y)\, f_Y(g^{-1}y)\, j(g^{-1}, y)\, \triangle(g^{-1})}{m(y)} \rho_l(dg)\, \mu_Y(dy) \\
&\quad = \int_{\mathsf{Y}} \left[\int_G \frac{h_2(gy)\, f_Y(gy)\, j(g, y)}{m(y)} \rho_l(dg) \right] h_1(y)\, m_\nu(y)\, \mu_Y(dy) \\
&\quad = \int_{\mathsf{Y}} (Rh_2)(y)\, h_1(y)\, m_\nu(y)\, \mu_Y(dy) \; ,
\end{aligned}
$$

where the third equality follows from (3.5), and the penultimate equality is due to (3.4). An application of Theorem 2.1 implies $k^* \succeq_E k_\nu$ and $\|K\| \leq \gamma^2 \|K_\nu\|$, where γ is the maximal correlation of the pair (Y, Y') whose joint distribution is (3.7). We now show that $\gamma = 1$. As pointed out in the proof of Theorem 2.1, γ^2 is the norm of the operator Q on $L^2_0(m_\nu)$ associated with the Mtf given by

$$
Q(y, dy') = \int_{y'' \in \mathsf{Y}} \tilde{R}(y'', dy')\, R(y, dy'') \; .
$$

We claim that $Q(y, dy') = \tilde{R}(y, dy')$ for all $y \in \mathsf{Y}$. Indeed, for $h \in L_0^2(m_\nu)$ and $y \in \mathsf{Y}$, we have

$$\begin{aligned}
(Qh)(y) &= \int_{\mathsf{Y}} (\tilde{R}h)(y'')\, R(y, dy'') \\
&= \int_G \frac{(\tilde{R}h)(gy)\, f_Y(gy)\, j(g, y)}{m(y)} \varrho_l(dg) \\
&= \int_G \int_G \frac{h(g'gy)\, m_\nu(g'gy)\, j(g', gy)}{m(gy)} \frac{f_Y(gy)\, j(g, y)}{m(y)} \varrho_l(dg')\, \varrho_l(dg) \\
&= \int_G \int_G \frac{h(g'gy)\, m_\nu(g'gy)\, j(g'g, y)\, f_Y(gy)}{m(gy)\, m(y)} \varrho_l(dg')\, \varrho_l(dg) \\
&= \int_G \int_G \frac{\triangle(g^{-1})\, h(g'y)\, m_\nu(g'y)\, j(g', y)\, f_Y(gy)}{m(gy)\, m(y)} \varrho_l(dg')\, \varrho_l(dg) \\
&= \left[\int_G \frac{h(g'y)\, m_\nu(g'y)\, j(g', y)}{m(y)} \varrho_l(dg') \right] \\
&\qquad \times \left[\int_G \frac{\triangle(g^{-1})\, f_Y(gy)\, j(g, y)}{\triangle(g^{-1})\, m(y)} \varrho_l(dg) \right] = (\tilde{R}h)(y) \,,
\end{aligned}$$

where the fifth equality follows from (3.3) and the penultimate equality is due to (3.5). Since $\tilde{R}$ is self-adjoint and idempotent, $\gamma^2 = \|Q\| = \|\tilde{R}\| = 1$ (see, e.g., Conway, 1990, Proposition 3.3). Hence, the result follows. □

Remark 3.1. Choi (2014) contains an alternative but more complicated proof of Proposition 3.1.

3.4. An illustration using Brownlee's stack loss data

We end this section with an illustration of the efficiency part of Proposition 3.1 by using a PX-DA algorithm and Choi and Hobert's (2013) Haar PX-DA algorithm for Bayesian linear regression with Laplace errors. To be specific, we will develop a PX-DA algorithm using the joint density upon which Choi and Hobert's (2013) DA algorithm is based and the group structure under which the Haar PX-DA algorithm is derived. We then compare the efficiency of the PX-DA and Haar PX-DA algorithms on Brownlee's (1965) stack loss dataset.

The Bayesian linear model with Laplace errors is formulated as follows. Let $\{Y_i\}_{i=1}^n$ be independent random variables such that

$$Y_i = x_i^T \beta + \sigma \epsilon_i \,,$$

where $x_i \in \mathbb{R}^p$ is a vector of known covariates associated with Y_i, $\beta \in \mathbb{R}^p$ is a vector of unknown regression coefficients, and $\sigma \in \mathbb{R}_+ := (0, \infty)$ is an unknown scale parameter. The errors, $\{\epsilon_i\}_{i=1}^\infty$, are assumed to be iid from the Laplace density with scale equal to two, so the common density is $d(\epsilon) := e^{-\frac{|\epsilon|}{2}}/4$. The standard default prior for (β, σ^2) is an improper prior that takes the form $\pi(\beta, \sigma^2) = (\sigma^2)^{-1} I_{\mathbb{R}_+}(\sigma^2)$. For inferential purposes, we would like to sample

from the posterior density of (β, σ^2). Let $y = (y_1, \ldots, y_n)^T$ be the vector of observed responses. The posterior density of (β, σ^2) is given by

$$f_{\beta,\sigma^2|Y}(\beta, \sigma^2 \,|\, y) = \frac{1}{c(y)} \frac{1}{4^n \sigma^n} \exp\left\{ -\frac{1}{2\sigma} \sum_{i=1}^{n} |y_i - x_i^T \beta| \right\} (\sigma^2)^{-1} I_{\mathbb{R}_+}(\sigma^2) \,, \tag{3.8}$$

where

$$c(y) = \int_{\mathbb{R}_+} \int_{\mathbb{R}^p} \frac{1}{4^n \sigma^n} \exp\left\{ -\frac{1}{2\sigma} \sum_{i=1}^{n} |y_i - x_i^T \beta| \right\} (\sigma^2)^{-1} d\beta \, d\sigma^2 \,.$$

As usual, let X denote the $n \times p$ matrix whose ith row is equal to x_i^T, and let $C(X)$ denote the column space of X. We assume throughout that X has full rank and $y \notin C(X)$ since these are necessary and sufficient conditions for $c(y) < \infty$ (Choi and Hobert, 2013, Proposition 1). A DA algorithm and the Haar PX-DA algorithm for exploring this intractable posterior are described in Choi and Hobert (2013) (hereafter, C&H). The DA algorithm is based on introducing a latent variable $Z = \{Z_i\}_{i=1}^n$, so the joint posterior density of β, σ^2 and Z, say $f(\beta, \sigma^2, z \,|\, y)$, is given by

$$\frac{1}{c(y)} \left[\prod_{i=1}^{n} \frac{z_i^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}} (\sigma^2)^{\frac{1}{2}}} \exp\left\{ -\frac{z_i}{2} \frac{(y_i - x_i^T \beta)^2}{\sigma^2} \right\} \right] \times \left[\prod_{i=1}^{n} \frac{1}{8 z_i^2} e^{-\frac{1}{8 z_i}} I_{\mathbb{R}_+}(z_i) \right] (\sigma^2)^{-1} I_{\mathbb{R}_+}(\sigma^2) \,. \tag{3.9}$$

A straightforward calculation using the well-known normal/inverse Gamma representation of the Laplace density shows that

$$\int_{\mathbb{R}_+^n} f(\beta, \sigma^2, z \,|\, y) \, dz = f_{\beta,\sigma^2|Y}(\beta, \sigma^2 \,|\, y) \,.$$

C&H's DA algorithm simply iterates between draws from the associated conditional densities, $f_{\beta,\sigma^2|Z,Y}(\beta, \sigma^2 \,|\, z, y)$ and $f_{Z|\beta,\sigma^2,Y}(z \,|\, \beta, \sigma^2, y)$, in the usual way. C&H show that the two conditionals can be specified as follows.

- The conditional distribution of (β, σ^2) is described as

  $$\beta \,|\, \sigma^2 = \tilde{\sigma}^2, Z = z, Y = y \sim N_p\left(\theta, \tilde{\sigma}^2 (X^T D^{-1} X)^{-1}\right) \,, \text{ and marginally,}$$

  $$\sigma^2 \,|\, Z = z, Y = y \sim \text{IG}\left(\frac{n-p}{2}, \frac{y^T D^{-1} y - \theta^T (X^T D^{-1} X) \theta}{2} \right) \,,$$

  where $\theta = (X^T D^{-1} X)^{-1} X^T D^{-1} y$ and D is a diagonal matrix whose ith diagonal element is z_i^{-1}. Also, when we write $W \sim \text{IG}(\alpha, \gamma)$, we mean that W has density proportional to

  $$w^{-\alpha-1} e^{-\gamma/w} I_{\mathbb{R}_+}(w) \,,$$

  where α and γ are strictly positive.

- $Z_1, \ldots, Z_n$ are conditionally independent with

$$Z_i \,|\, \beta = \tilde{\beta}, \sigma^2 = \tilde{\sigma}^2, Y = y \sim \begin{cases} \text{Inv Gau}\left(\frac{\tilde{\sigma}}{2|y_i - x_i^T\tilde{\beta}|}, \frac{1}{4}\right) & \text{if } |y_i - x_i^T\tilde{\beta}| > 0 \\ \text{IG}\left(\frac{1}{2}, \frac{1}{8}\right) & \text{if } |y_i - x_i^T\tilde{\beta}| = 0 \end{cases}$$

  Here, when we write $W \sim \text{Inv Gau}(\mu, \lambda)$, we mean that W has density given by

$$\sqrt{\frac{\lambda}{2\pi w^3}} \exp\left\{-\frac{\lambda(w-\mu)^2}{2\mu^2 w}\right\} I_{\mathbb{R}_+}(w) \,,$$

  where μ and λ are strictly positive.

C&H's Haar PX-DA algorithm is derived under the group $G = \mathbb{R}_+$, which acts on $\mathbb{R}_+^n$ (the space of the latent variable Z) through scalar multiplication. In particular, the sandwich step is formed using the recipe described in (3.6) with $j(g, z) = g^n$ and $\varrho_l(dg) = dg/g$, and f_Y equal to the z-marginal density of (3.9). It is shown that the sandwich step corresponds to the move $z \to z' = gz$ with g drawn from $\text{IG}(n, \sum_{i=1}^n \frac{1}{8z_i})$ distribution.

We now develop a PX-DA algorithm using (3.9) and the group structure on G. Consider a conditional probability measure, $I(\sigma^2, dg)$, on G given (β, σ^2) that depends on σ^2 but not on β and has a point mass at σ^2. (Note that σ^2 lives on $\mathbb{R}_+$, so $I(\sigma^2, dg)$ is legitimate.) As described in (3.1), define a probability density such that

$$\tilde{f}^{(I)}(\beta, \sigma^2, z \,|\, y) = \int_G f(\beta, \sigma^2, gz \,|\, y)\, g^n\, I(\sigma^2, dg) \,.$$

A straightforward manipulation reveals that $\tilde{f}^{(I)}(\beta, \sigma^2, z \,|\, y)$ is equal to

$$\frac{1}{c(y)} \left[\prod_{i=1}^n \left(\frac{2\pi}{z_i}\right)^{-\frac{1}{2}} \exp\left\{-\frac{z_i}{2}(y_i - x_i^T\beta)^2\right\}\right] \times \left[\prod_{i=1}^n \frac{1}{8\sigma^2} z_i^{-2} \exp\left\{-\frac{1}{8\sigma^2 z_i}\right\} I_{\mathbb{R}_+}(z_i)\right] (\sigma^2)^{-1} I_{\mathbb{R}_+}(\sigma^2) \,.$$

By construction, the (β, σ^2)-marginal density of $\tilde{f}^{(I)}(\beta, \sigma^2, z \,|\, y)$ is the target (3.8), and the DA algorithm based on the new joint density $\tilde{f}^{(I)}(\beta, \sigma^2, z \,|\, y)$ is a PX-DA algorithm. The associated conditional densities, $\tilde{f}^{(I)}_{\beta,\sigma^2|Z,Y}(\beta, \sigma^2 \,|\, z, y)$ and $\tilde{f}^{(I)}_{Z|\beta,\sigma^2,Y}(z \,|\, \beta, \sigma^2, y)$, to simulate the PX-DA algorithm can be easily derived using the similar arguments in Section 2 of C&H, along with the conditional independence of β and σ^2 given (z, y) as follows.

- β and σ^2 are conditionally independent with

$$\beta \,|\, Z = z, Y = y \sim N_p\left(\theta, (X^T D^{-1} X)^{-1}\right) \,, \text{ and}$$

$$\sigma^2 \,|\, Z = z, Y = y \sim \text{IG}\left(n, \sum_{i=1}^n \frac{1}{8z_i}\right)$$

TABLE 1
Results based on 10^5 *iterations*

Parameter	Results for the PX-DA algorithm		Results for C&H's Haar PX-DA algorithm	
σ^2	Estimate	Standard error	Estimate	Standard Error
	2.054	0.008703	2.064	0.004173

- $Z_1, \ldots, Z_n$ are conditionally independent with

$$Z_i \,|\, \beta = \tilde{\beta}, \sigma^2 = \tilde{\sigma}^2, Y = y \sim \begin{cases} \text{Inv Gau}\left(\frac{\tilde{\sigma}^{-1}}{2|y_i - x_i^T\tilde{\beta}|}, \frac{1}{4\tilde{\sigma}^2}\right) & \text{if } |y_i - x_i^T\tilde{\beta}| > 0 \\ \text{IG}\left(\frac{1}{2}, \frac{1}{8\tilde{\sigma}^2}\right) & \text{if } |y_i - x_i^T\tilde{\beta}| = 0 \end{cases}$$

We implement the PX-DA algorithm and C&H's Haar PX-DA algorithm on Brownlee's (1965) stack loss data. The data are from the operation of a plant for the oxidation of ammonia to nitric acid, measured on 21 consecutive days. The dataset consists of $\{(y_i, x_{i1})\}_{i=1}^{21}$, where x_{i1} is a covariate indicating the air flow to the plant, and y_i is the percentage of ammonia lost (times 10). We consider the model

$$Y_i = \beta_0 + \beta_1 x_{i1} + \sigma\epsilon_i \ ,$$

where ϵ_i's are iid with the common Laplace density $d(\epsilon)$, with the standard default prior $\pi(\beta, \sigma^2)$. It can be easily verified (using C&H's necessary and sufficient conditions for posterior propriety described above) that the posterior density is proper. For all Markov chains, we choose the initial value of σ^2 to be 1 and the initial value of β to be the maximum likelihood estimate under the standard linear model (with Gaussian errors).

We run the PX-DA and C&H's Haar PX-DA algorithms for a burn-in period of 4×10^5 iterations. The next 10^5 iterations are used to obtain the posterior expectation for the two Markov chains, and we adopt the batch means method to estimate the asymptotic variances. (See Jones et al. (2006) for precise formula and theoretical properties of the batch means estimator.) For this particular example, we are interested in estimating the posterior expectation of $h(\beta, \sigma^2) = \sigma^2$ as Yu and Moyeed (2001) fit similar Bayesian regression models with fixed scale parameter ($\sigma^2 = 1$). Also, similar arguments to the proof of C&H's Proposition 1 imply that, in the current setting, $|h(\beta, \sigma^2)|^3$ is integrable with respect to the posterior. Table 1 provides the simulation results. Note that the estimated asymptotic standard error for the PX-DA algorithm is 2.09 times as large as the corresponding value for C&H's Haar PX-DA algorithm. This suggests that, in this particular example, the PX-DA algorithm requires about $2.09^2 = 4.35$ times as many iterations as C&H's PX-DA algorithm to achieve the same level of precision.

4. Improving the collapsing theorem

Let $\mathsf{X}, \mathsf{Y}, \mu_X$ and μ_Y be as in Section 2.2. Let Z be a separable metric space equipped with its Borel σ-algebra, and assume that Z is sub-Cauchy. Assume

also that μ_Z is a σ-finite measure on Z. As before, suppose f_X is an intractable density with respect to μ_X on X that we would like to explore. Let $f : \mathsf{X} \times \mathsf{Y} \times \mathsf{Z} \to [0, \infty)$ be a joint density with respect to $\mu_X \times \mu_Y \times \mu_Z$ such that $\int_{\mathsf{Y}} \int_{\mathsf{Z}} f(x, y, z)\, \mu_Z(dz)\, \mu_Y(dy) = f_X(x)$. As usual, let $f_{X|Y,Z}, f_{Y,Z|X}, f_{X|Y}$ and $f_{Y|X}$ be conditional densities associated with f. Consider two DA chains based on Mtds k and $\tilde{k}$ given by

$$k(x \,|\, x') = \int_{\mathsf{Y}} \int_{\mathsf{Z}} f_{X|Y,Z}(x \,|\, y, z)\, f_{Y,Z|X}(y, z \,|\, x')\, \mu_Z(dz)\, \mu_Y(dy) \,,$$
$$\tilde{k}(x \,|\, x') = \int_{\mathsf{Y}} f_{X|Y}(x \,|\, y)\, f_{Y|X}(y \,|\, x')\, \mu_Y(dy) \,.$$

Denote the operators on $L_0^2(f_X)$ corresponding to k and $\tilde{k}$ by K and $\tilde{K}$. It follows from Liu et al. (1994) and Liu (1994) that $\|\tilde{K}\| \leq \|K\|$, which is called the collapsing theorem. Here, we use Theorem 2.1 to improve the collapsing theorem in terms of efficiency ordering.

Proposition 4.1. *Assume k and $\tilde{k}$ are Harris ergodic. Then $\|\tilde{K}\| \leq \|K\|$ and $\tilde{k} \succeq_E k$.*

Proof. Let π be an arbitrary density on Z with respect to μ_Z, and let $f_{X,Y}$ denote the marginal density associated with f. It is easy to see that $\tilde{k}$ can be written as the DA Mtd based on the joint density $f^*(x, y, z) = f_{X,Y}(x, y)\, \pi(z)$, which is simulated by drawing alternately from the associated conditional densities $f^*_{Y,Z|X}$ and $f^*_{X|Y,Z}$. Indeed, $f^*_{X|Y,Z}(x \,|\, y, z) = f_{X|Y}(x \,|\, y)$ and $f^*_{Y,Z|X}(y, z \,|\, x) = f_{Y|X}(y \,|\, x)\, \pi(z)$, so we have

$$\int_{\mathsf{Y}} \int_{\mathsf{Z}} f^*_{X|Y,Z}(x \,|\, y, z)\, f^*_{Y,Z|X}(y, z \,|\, x')\, \mu_Z(dz)\, \mu_Y(dy)$$
$$= \int_{\mathsf{Y}} \int_{\mathsf{Z}} f_{X|Y}(x \,|\, y)\, f_{Y|X}(y \,|\, x')\, \pi(z)\, \mu_Z(dz)\, \mu_Y(dy) = \tilde{k}(x \,|\, x') \,.$$

We now establish the two conditions of Theorem 2.1. Let R be a Mtf on $\mathsf{Y} \times \mathsf{Z}$ defined by

$$R\big((y, z), (dy' \times dz')\big) = I(y, dy')\, f_{Z|Y}(z' \,|\, y')\, \mu_Z(dz') \,,$$

where $I(y, dy')$ is the trivial Mtf that is a point mass at y, and $f_{Z|Y}$ is the conditional density associated with f. Then

$$\int_{\mathsf{Y}} \int_{\mathsf{Z}} f_{X|Y,Z}(x \,|\, y', z')\, R\big((y, z), (dy' \times dz')\big)$$
$$= \int_{\mathsf{Y}} \left[\int_{\mathsf{Z}} f_{X|Y,Z}(x \,|\, y', z')\, f_{Z|Y}(z' \,|\, y')\, \mu_Z(dz') \right] I(y, dy')$$
$$= \int_{\mathsf{Y}} f_{X|Y}(x \,|\, y')\, I(y, dy') = f_{X|Y}(x \,|\, y) = f^*_{X|Y,Z}(x \,|\, y, z) \,.$$

Since $f^*_{Y,Z}(y,z) = \int_{\mathsf{X}} f^*(x,y,z)\,\mu_X(dx) = f_Y(y)\,\pi(z)$, we have

$$\begin{aligned}
&\int_{y\in\mathsf{Y}}\int_{z\in\mathsf{Z}} R\big((y,z),(dy'\times dz')\big)\, f^*_{Y,Z}(y,z)\,\mu_Z(dz)\,\mu_Y(dy)\\
&\qquad = f_{Z|Y}(z'\,|\,y')\,\mu_Z(dz')\int_{y\in\mathsf{Y}}\left[\int_{\mathsf{Z}}\pi(z)\,\mu_Z(dz)\right] I(y,dy')\,f_Y(y)\,\mu_Y(dy)\\
&\qquad = f_{Z|Y}(z'\,|\,y')\,f_Y(y')\,\mu_Y(dy')\,\mu_Z(dz')\\
&\qquad = f_{Y,Z}(y',z')\,\mu_Y(dy')\,\mu_Z(dz')\ .
\end{aligned}$$

An application of Theorem 2.1 implies $\tilde{k} \succeq_E k$ and $\|\tilde{K}\| \le \gamma^2\|K\|$, where γ is the maximal correlation of the random pair $((Y,Z),(Y',Z'))$, whose joint distribution is

$$R\big((y,z),(dy'\times dz')\big)\, f^*_{Y,Z}(y,z)\,\mu_Y(dy)\,\mu_Z(dz)\ .$$

We now show that $\gamma = 1$. It suffices to establish that, for some function $g(y,z)$ on $\mathsf{Y}\times\mathsf{Z}$ such that $0 < \mathrm{Var}\{g(Y',Z')\} < \infty$,

$$\mathrm{Var}\left[\mathrm{E}\left\{g(Y',Z')\,|\,Y,Z\right\}\right] = \mathrm{Var}\left\{g(Y',Z')\right\}\ .$$

Take an arbitrary nonzero function $h \in L^2_0(f_Y)$, and define a function $g(y,z)$ on $\mathsf{Y}\times\mathsf{Z}$ as $g(y,z) = h(y)$. It is easy to see that $0 < \mathrm{Var}\{g(Y',Z')\} < \infty$ and that, for all $(y,z) \in \mathsf{Y}\times\mathsf{Z}$,

$$\begin{aligned}
&\mathrm{E}\left\{g(Y',Z')\,|\,(Y,Z) = (y,z)\right\}\\
&\quad = \int_{\mathsf{Y}}\int_{\mathsf{Z}} h(y')\,R\big((y,z),(dy'\times dz')\big)\\
&\quad = \int_{\mathsf{Y}} h(y')\left[\int_{\mathsf{Z}} f_{Z|Y}(z'\,|\,y')\,\mu_Z(dz')\right] I(y,dy') = h(y)\ .
\end{aligned}$$

So, we have

$$\begin{aligned}
\mathrm{Var}\left[\mathrm{E}\left\{g(Y',Z')\,|\,Y,Z\right\}\right] &= \int_{\mathsf{Y}}\int_{\mathsf{Z}} h^2(y)\, f^*_{Y,Z}(y,z)\,\mu_Z(dz)\,\mu_Y(dy)\\
&= \int_{\mathsf{Y}} h^2(y)\, f_Y(y)\int_{\mathsf{Z}}\pi(z)\,\mu_Z(dz)\,\mu_Y(dy)\\
&= \int_{\mathsf{Y}} h^2(y)\, f_Y(y)\,\mu_Y(dy)\\
&= \mathrm{Var}\left\{h(Y')\right\} = \mathrm{Var}\left\{g(Y',Z')\right\}\ ,
\end{aligned}$$

which completes the proof. □

Acknowledgements

The second author was supported by NSF Grants DMS-11-06395 and DMS-15-11945

References

Brownlee, K. A. (1965). *Statistical Theory and Methodology in Science and Engineering*. 2nd ed. Wiley, New York. MR0119268

Choi, H. M. (2014). *Convergence Analysis of Gibbs Samplers for Bayesian Regression Models*. Ph.D. thesis, University of Florida. MR3439027

Choi, H. M. and Hobert, J. P. (2013). Analysis of MCMC algorithms for Bayesian linear regression with Laplace errors. *Journal of Multivariate Analysis*, **117** 32–40. MR3053533

Conway, J. B. (1990). *A Course in Functional Analysis*. 2nd ed. Springer, New York. MR1070713

Faden, A. M. (1985). The existence of regular conditional probabilities: Necessary and sufficient conditions. *The Annals of Probability*, **13** 288–298. MR0770643

Gebelein, H. (1941). Das statistische Problem der Korrelation als Variations und Eigenwertproblem und sein Zusammenhang mit der Ausgleichsrechnung. *Zeitschrift für Angewandte Mathematik und Mechanik*, **21** 364–379. MR0007220

Hobert, J. P. and Marchev, D. (2008). A theoretical comparison of the data augmentation, marginal augmentation and PX-DA algorithms. *The Annals of Statistics*, **36** 532–554. MR2396806

Hobert, J. P. and Román, J. C. (2011). Comment: "To center or not to center: That is not the question – An Ancillarity-Sufficiency Interweaving Strategy (ASIS) for boosting MCMC efficiency". *Journal of Computational and Graphical Statistics*, **20** 571–580. MR2878988

Jones, G. L., Haran, M., Caffo, B. S. and Neath, R. (2006). Fixed-width output analysis for Markov chain Monte Carlo. *Journal of the American Statistical Association*, **101** 1537–1547. MR2279478

Khare, K. and Hobert, J. P. (2011). A spectral analytic comparison of trace-class data augmentation algorithms and their sandwich variants. *The Annals of Statistics*, **39** 2585–2606. MR2906879

Lancaster, H. O. (1957). Some properties of the bivariate normal distribution considered in the form of a contingency table. *Biometrika*, **44** 289–292.

Liu, J. S. (1994). The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem. *Journal of the American Statistical Association*, **89** 958–966. MR1294740

Liu, J. S., Wong, W. H. and Kong, A. (1994). Covariance structure of the Gibbs sampler with applications to comparisons of estimators and augmentation schemes. *Biometrika*, **81** 27–40. MR1279653

Liu, J. S., Wong, W. H. and Kong, A. (1995). Covariance structure and convergence rate of the Gibbs sampler with various scans. *Journal of the Royal Statistical Society, Series B*, **57** 157–169. MR1325382

Liu, J. S. and Wu, Y. N. (1999). Parameter expansion for data augmentation. *Journal of the American Statistical Association*, **94** 1264–1274. MR1731488

Mira, A. and Geyer, C. J. (1999). Ordering Monte Carlo Markov chains. Tech. rep., School of Statistics, University of Minnesota.

Parthasarathy, K. R. (1967). *Probability Measures on Metric Spaces*. Academic Press, New York. MR0226684

Ramachandran, D. (1979). *Perfect Measures: Basic theory*. ISI lecture notes, Macmillian. MR0553600

Roberts, G. O. and Rosenthal, J. S. (1997). Geometric ergodicity and hybrid Markov chains. *Electronic Communications in Probability*, **2** 13–25. MR1448322

Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation (with discussion). *Journal of the American Statistical Association*, **82** 528–550. MR0898357

Yu, K. and Moyeed, R. A. (2001). Bayesian quantile regression. *Statistics & Probability Letters*, **54** 437–447. MR1861390

Yu, Y. and Meng, X. L. (2011). To center or not to center: That is not the question – An Ancillarity-Sufficiency Interweaving Strategy (ASIS) for boosting MCMC efficiency (with discussion). *Journal of Computational and Graphical Statistics*, **20** 531–615. MR2878987