Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Econometrics 123 (2004) 201–225
JOURNAL OF Econometrics
www.elsevier.com/locate/econbase

Adaptive radial-based direction sampling: some flexible and robust Monte Carlo integration methods

Luc Bauwens[a], Charles S. Bos[b], Herman K. van Dijk[c,*], Rutger D. van Oest[c]

[a] *CORE and Department of Economics, Université catholique de Louvain, Belgium*
[b] *Department of Econometrics & O.R., Free University Amsterdam, The Netherlands*
[c] *Econometric and Tinbergen Institutes, Erasmus University Rotterdam, The Netherlands*

Abstract

Adaptive radial-based direction sampling (ARDS) algorithms are specified for Bayesian analysis of models with non-elliptical, possibly, multimodal target distributions. A key step is a radial-based transformation to directions and distances. After the transformation a Metropolis-Hastings method or, alternatively, an importance sampling method is applied to evaluate generated directions. Next, distances are generated from the exact target distribution. An adaptive procedure is applied to update the initial location and covariance matrix in order to sample directions in an efficient way. The ARDS algorithms are illustrated on a regression model with scale contamination and a mixture model for economic growth of the USA.

JEL classification: C11; C15; C63

Keywords: Markov chain Monte Carlo; Importance sampling; Radial coordinates

1. Introduction

In recent decades Markov Chain Monte Carlo (MCMC) methods, in particular Metropolis-Hastings (MH) and Gibbs sampling (GS) and, to a lesser extent, independent sampling methods like importance sampling (IS), have been applied

* Corresponding author. Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands. Tel.: +3110-4082382; fax: +3110-4089162.
E-mail addresses: bauwens@core.ucl.ac.be (L. Bauwens), cbos@feweb.vu.nl (C.S. Bos), hkvandijk@few.eur.nl (H.K. van Dijk), vanoest@few.eur.nl (R.D. van Oest).

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2003.12.002

extensively and successfully within Bayesian analyses of statistical and econometric models.

Although Monte Carlo (MC) methods revolutionized the applicability of Bayesian inference, there is, in practice, a substantial variation in their convergence behaviour. The special features of the sampling method, the complex structure of the model, or the nature of the data may be the culprit of such behaviour. Hobert and Casella (1996) show for instance that the Gibbs sampler does not converge for the case of a hierarchical linear mixed model when the prior is uniform. Other examples of complex models are the ones with reduced rank structures. Kleibergen and Van Dijk (1994, 1998) demonstrate near reducibility of MCMC methods when there exists near non-identifiability and non-stationarity in econometric models with flat priors. Justel and Peña (1996) emphasize the convergence problems of the Gibbs sampler when there are outliers in the data. The performance of the Gibbs sampler is also seriously hampered by strong correlation in the target distribution. Convergence problems of importance sampling using a simple normal or Student t candidate density have been documented by Van Dijk and Kloek (1984) and Geweke (1989). A multimodal target density may pose problems to all methods. If the MH candidate density is unimodal, with low probability of drawing candidate values in one of the modes, this mode may be missed completely, even when the sample size is large. More generally stated, the acceptance probability may be very low, as many candidate values lying between the modes have to be rejected. With the Gibbs sampler, reducibility of the chain may occur in this case. Using a unimodal normal or Student t candidate function, the method of importance sampling ends up with many drawings having only negligible weights. A common difficulty encountered in all samplers is the choice of a candidate or importance density when little is known about the shape of the target density. In such a case, updating the candidate density sequentially is a partial solution. [1]

In this paper we introduce the class of adaptive radial-based direction sampling (ARDS) methods to sample from a target (posterior) distribution which is possibly multi-modal, skew, and exhibits strong correlation, in summary is non-elliptical. The ARDS algorithms feature a composite transformation. The key step is one where the m-dimensional parameter space is transformed into radial coordinates which consist of a distance measure and a $(m-1)$-dimensional vector of directions. A MH or an IS algorithm is applied to sample the $(m-1)$ directions. Next, a distance is sampled conditionally on the directions, from the (transformed) exact target density, by the inverse transformation method. A location-scale transformation is used as part of the transformation to radial coordinates and this transformation is sequentially updated, using the posterior first and second order moments obtained in successive rounds of the algorithm. The adaptive procedure is intended to improve the acceptance rate within the MH step and to give a more uniform distribution of the weights in the IS step.

[1] This corresponds to the experimental results obtained by local adaptive importance sampling when the posterior is ill behaved, see e.g. Van Dijk and Kloek (1980), Oh and Berger (1992), and Givens and Raftery (1996).

The advantages of the ARDS algorithms are threefold. Firstly, the algorithms are *parsimonious* in their use of information on the shape of the target density. Only location and scale need to be specified as initial values. Secondly, the algorithms are *flexible* and *robust*: they can handle a large variety of features of target distributions, in particular multimodality, strong correlation, extreme skewness, and heavy tails. The ARDS algorithms avoid the often time-consuming and difficult task, especially for non-experts, of choosing and tuning a sampling algorithm for a specific application, such as computation of posterior moments in Bayesian inference. They can be applied without the need to study in depth the shape of the posterior density in order to design a sophisticated approximation to it. For any specific model and data combination, a more efficient algorithm than ARDS may be designed, but our viewpoint is that the extra effort (in research time) required to achieve this may not be rewarding. Furthermore, we point out that one may apply ARDS algorithms as a first step to explore the shape of the posterior. Given knowledge of this shape one may design a specific algorithm more tailored to the problem studied. Thirdly, the algorithms can handle multiple *linear inequality conditions* on the parameter space without any additional complications for the implementation. In practice researchers often make use of bounds on parameters.

The ARDS algorithms extend earlier methods like the method of Box and Muller (1958), the adaptive direction sampling (ADS) algorithms proposed by Gilks et al. (1994), the mixed integration method by Van Dijk et al. (1985), and the spherical integration method by Monahan and Genz (1997). Details are given in Section 2.4.

The outline of the paper is as follows. In Section 2 the algorithms are introduced. In Section 3 some models are used for experimenting with ARDS and for comparing its performance with that of the standard Metropolis-Hastings and importance sampling algorithms. The models feature multimodality, high correlation and skewness. As empirical examples we use a regression model with scale contamination in order to investigate a study from Justel and Peña (1996) concerning the oxidation of ammonia to nitric acid in a plant. Next, we analyze economic growth of the USA using a mixture model as in Frühwirth-Schnatter (2001). Conclusions are presented in Section 4.

2. Adaptive radial-based direction sampling

Most simulation algorithms for posterior distributions generate random drawings in the original parameter space. Several researchers advocate to simulate in a transformed space, where the simulation is more efficient in some sense, see e.g. Gilks and Roberts (1996). For example, if there exists a strong correlation between two random variables, an orthogonalising transformation reduces serial dependence in a Gibbs sampling scheme. Another example arises in the context of importance sampling: an efficient importance function is easier to construct when an adequate transformation yields a distribution that is much closer to a symmetric one than the original one. The adaptive radial-based sampling algorithms rely on this general idea. They are based on a composite transformation to radial coordinates. Heuristically, the original parameter space is transformed into a $(m-1)$-dimensional space of directions and a unidimensional complementary space of distances. When the target or candidate density is a member

of the elliptical family one can use the following result. The transformed density of the $(m-1)$-dimensional directions is uniform on the unit sphere and this density is independent of the density of the unidimensional distance which has a known analytical form (e.g. a member of the gamma family when the candidate is normal); see e.g. Muirhead (1982, Section 1.5). This result is the basis for the Box–Muller method of generating normal random variables; see Box and Muller (1958). In our approach, one key step is the creation of directions, or lines, where the target density has non-trivial mass. A second key step is generating distances, or points on the lines, from the exact target density by means of the numerical inverse transformation method.

In this section, we concentrate on two members of the radial-based algorithms, one is of the Metropolis-Hastings type, and the second one is of the importance sampling type. These variants will be referred to as adaptive radial-based Metropolis-Hastings sampling (ARMHS) and adaptive radial-based importance sampling (ARIS), respectively.

2.1. The radial transformation

Since the radial transformation is at the heart of the algorithms, we start by describing the transformation from Cartesian coordinates to radial coordinates. In our notation, $y=(y_1,\ldots,y_m)$ denotes the Cartesian coordinates of a point, and (ρ,η) denotes the corresponding radial coordinates. Here $\eta=(\eta_1,\ldots,\eta_{m-1})$ indicates the direction of the point relative to the origin, and ρ is related to the Euclidean distance.

In a general form, the radial transformation may be characterized by its inverse transformation

$$\begin{aligned} y_1 &= \rho h_1(\eta_1,\ldots,\eta_{m-1}),\\ &\vdots\\ y_{m-1} &= \rho h_{m-1}(\eta_1,\ldots,\eta_{m-1}),\\ y_m &= \rho h_m(\eta_1,\ldots,\eta_{m-1}), \end{aligned} \tag{1}$$

where $m \geqslant 2$ is the dimension, and $h(\eta)=(h_1(\eta),\ldots,h_m(\eta))$ are differentiable functions. The radial transformation maps the Cartesian coordinates y to a position $h(\eta)$ on the unit circle (the direction) and a stretching factor ρ which further determines the position of y given $h(\eta)$ (the distance). It can be shown that the Jacobian of the general radial transformation is

$$\begin{aligned} J_y(\rho,\eta) &= \det\begin{pmatrix} \dfrac{\partial y(\rho,\eta)}{\partial \eta'} & \dfrac{\partial y(\rho,\eta)}{\partial \rho} \end{pmatrix} = \rho^{m-1}\det\begin{pmatrix} \dfrac{\partial h(\eta)}{\partial \eta'} & h(\eta) \end{pmatrix}\\ &\equiv J_y(\rho)J_y(\eta). \end{aligned} \tag{2}$$

The form of this Jacobian has some important implications which are used in ARDS. It turns out that implementation of the ARDS algorithms is only based on the Jacobian factor $J_y(\rho)$, and does not depend on $J_y(\eta)$, see Proposition 1 in Section 2.2. The Jacobian factor $J_y(\eta)$ drops out of the calculations, because of the multiplicative

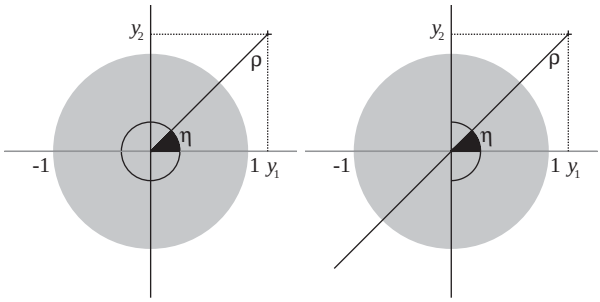

Fig. 1. The relationship between Cartesian coordinates and polar coordinates in the two-dimensional case: standard polar coordinates in the left panel and signed polar coordinates in the right panel.

structure of (2). As $J_y(\rho) = \rho^{m-1}$ is invariant with respect to the functions $h_1, \ldots, h_m$, our approach can be applied to any transformation satisfying (1).

2.1.1. The polar transformation

A well known special case of the general transformation described above is the two-dimensional polar transformation with inverse transformation

$$y_1 = \rho \cos(\eta), \tag{3}$$

$$y_2 = \rho \sin(\eta). \tag{4}$$

The standard polar transformation from $y \in \mathbb{R}^2$ to $(\rho, \eta) \in \mathbb{R}^+ \times (0, 2\pi)$, which is one-to-one with (3)–(4), is given by

$$\rho = \sqrt{y_1^2 + y_2^2}, \tag{5}$$

$$\eta = \operatorname{sgn}(y_2) \arccos(y_1/\rho). \tag{6}$$

The left panel of Fig. 1 illustrates the relationship between orthogonal coordinates and standard polar coordinates. A feature of the standard polar transformation is that $\rho \in R^+$, implying that the direction η only defines one half of a line, and not an entire line. However, we recall that the second key step of ARDS consists of sampling from the exact target density on *entire* lines, so that information from the *whole* target density is considered given a direction η. The signed polar transformation from $y \in \mathbb{R}^2$ to $(\rho, \eta) \in \mathbb{R} \times (-\pi/2, \pi/2)$ is a variant of the standard polar transformation such that η does define a whole line. It is given by

$$\rho = \operatorname{sgn}(y_1) \sqrt{y_1^2 + y_2^2}, \tag{7}$$

$$\eta = \arcsin(y_2/\rho) \tag{8}$$

with inverse transformation again defined by (3)–(4), see Bauwens et al. (2002). The relationship between orthogonal coordinates and signed polar coordinates is shown in the right panel of Fig. 1.

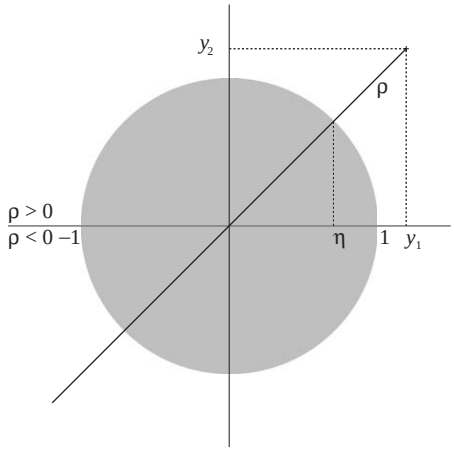

Fig. 2. The relationship between Cartesian coordinates and radial coordinates in the two-dimensional case.

2.1.2. *An efficient radial transformation*

One drawback of the polar transformation is that it is possible but not straightforward to generalize it to more than two dimensions, see Bauwens et al. (2002) and Muirhead (1982, Theorems 1.5.5 and 2.1.3). Furthermore, the polar transformation is computationally not very efficient. We therefore propose a transformation which satisfies (1), is easy to generalize to more than two dimensions, and is more efficient than the polar transformation. For m dimensions, the transformation from $(y_1,\ldots,y_m)\in\mathbb{R}^m$ to $(\rho,\eta)=(\rho,\eta_1,\ldots,\eta_{m-1})\in\mathbb{R}\times\{\eta\in\mathbb{R}^{m-1}:\eta'\eta<1\}$ is given by

$$\rho=\operatorname{sgn}(y_m)\sqrt{y'y}, \tag{9}$$

$$\eta_j=\frac{y_j}{\rho},\quad j=1,\ldots,m-1, \tag{10}$$

with inverse transformation

$$y_j=\rho\eta_j,\quad j=1,\ldots,m-1, \tag{11}$$

$$y_m=\rho\sqrt{1-\eta'\eta}. \tag{12}$$

The Jacobian of this transformation is

$$J_y(\rho,\eta)=J_y(\rho)J_y(\eta)=\rho^{m-1}(1-\eta'\eta)^{-1/2}. \tag{13}$$

Basically, y is transformed to $m-1$ Cartesian coordinates η on the unit circle and a stretching factor ρ. This is illustrated in Fig. 2 for $m=2$ dimensions. Here we note that the sign of ρ determines whether y is located above or below the y_1 axis.

2.2. Adaptive radial-based Metropolis-Hastings sampling

We start by defining the radial-based Metropolis-Hastings algorithm (RMHS), which is based on a candidate generating density that is taken to be multivariate normal with parameters μ and Σ. However, we will show that actually any elliptically contoured candidate distribution can be considered without affecting the sampling results, see Proposition 1 and Remark 1. Any density within the elliptical class is of the form $c_m(\det(\Sigma))^{-1/2}g((x-\mu)'\Sigma^{-1}(x-\mu))$, where c_m is some normalizing constant, see e.g. Muirhead (1982, Section 1.5). Elliptical distributions are symmetric, unimodal, and their kurtosis is determined by the function g. A fundamental property is that they are preserved under affine transformations, although the mean and covariance matrix might change. This property is essential for the derivation of the ARDS algorithms, see (A.3) in Appendix A. The normal density arises when $g(u)=\exp(-u/2)$, whereas, for example, the m-variate Student-t density with v degrees of freedom corresponds to $g(u)=(1+u/v)^{-(v+m)/2}$. As drawing from a normal distribution is more efficient than drawing from a Student t density or any other elliptically contoured density, the normal distribution is used in the implementation of RMHS and any other ARDS algorithm. We emphasize that, although for expository purpose the subsequent discussion focusses on convenient normal candidate drawings, these drawings may be generated from any member of the class of elliptical distributions. After defining RMHS, we continue this subsection by illustrating the algorithm for a bimodal target density. Finally, we define the adaptive RMHS algorithm (ARMHS), where μ and Σ are updated using the sample of draws from a previous round of the RMHS algorithm.

2.2.1. Definition of radial-based Metropolis-Hastings sampling

RMHS is based on an independence chain MH algorithm. It uses draws from a $N(\mu,\Sigma)$ candidate where hopefully μ and Σ provide good approximations to the unknown mean and covariance matrix of the target distribution, see below. In contrast with the MH algorithm, the drawings are not used for construction of a Markov chain in the original parameter space. Instead, a composite transformation is made. For expository purpose we treat this transformation explicitly in two steps.

The first step is a location-scale transformation of a realization x to a realization y. This transformation aims at standardizing the candidate density with respect to the location, scale, and correlations of the target (posterior) density, denoted by $p(x)$. It is defined by the affine transformation[2]

$$y = y(x|\mu,\Sigma) = \Sigma^{-1/2}(x-\mu), \tag{14}$$

with inverse transformation

$$x = x(y|\mu,\Sigma) = \mu + \Sigma^{1/2}y, \tag{15}$$

and Jacobian

$$J_x(y) = \det(\Sigma^{1/2}). \tag{16}$$

[2] $\Sigma^{1/2}$ denotes the Cholesky decomposition of Σ, and $\Sigma^{-1/2}$ denotes the inverse matrix of $\Sigma^{1/2}$.

Table 1
One iteration of RMHS

1. Generate x_i^* from $N(\mu, \Sigma)$
2. Transform x_i^* to $y_i^* = \Sigma^{-1/2}(x_i^* - \mu)$
3. Transform y_i^* to ρ_i^* and η_i^*, using (9) and (10)
4. Apply MH step to η_i, see (20)
5. Generate ρ_i from $p(\rho\|\eta_i)$ by inverting numerically its cdf
6. Transform ρ_i and η_i to y_i, using (11) and (12)
7. Transform y_i to $x_i = \mu + \Sigma^{1/2} y_i$

Note that steps 1 and 2 amount to generating y_i^* from $N(0, I_m)$. We want to make explicit the dependence on μ and Σ.

The second key step is the radial transformation, which is defined by (9) and (10), with inverse transformation given by (11) and (12), and Jacobian (13).

Combining the two transformations, one obtains the composite transformation

$$\begin{pmatrix} \rho \\ \eta \end{pmatrix} = \begin{pmatrix} \rho(x|\mu, \Sigma) \\ \eta(x|\mu, \Sigma) \end{pmatrix} = \begin{pmatrix} \rho(y(x|\mu, \Sigma)) \\ \eta(y(x|\mu, \Sigma)) \end{pmatrix} \tag{17}$$

with inverse transformation

$$x = x(\rho, \eta|\mu, \Sigma) = x(y(\rho, \eta)|\mu, \Sigma) \tag{18}$$

and Jacobian

$$J_x(\rho, \eta) = J_y(\rho, \eta) J_x(y) = J_y(\rho) J_y(\eta) \det(\Sigma^{1/2}). \tag{19}$$

Applying the two transformations to a candidate realization x_i^* from $N(\mu, \Sigma)$ yields a distance ρ_i^* and a vector of directions η_i^* (hereafter referred to as a 'direction').[3] Ignoring the distance, the candidate direction is either accepted or rejected in an MH step, that is, the direction becomes

$$\eta_i = \begin{cases} \eta_i^* & \text{with probability } \alpha(\eta_{i-1}, \eta_i^*), \\ \eta_{i-1} & \text{with probability } 1 - \alpha(\eta_{i-1}, \eta_i^*) \end{cases} \tag{20}$$

for some acceptance probability $\alpha(\eta_{i-1}, \eta_i^*)$, which is given in Proposition 1 below. An iteration of RMHS is completed by drawing from the target distribution on the line defined by the direction η_i. This can be done as follows. First, one draws a distance ρ_i from the transformed target distribution for given direction η_i using the numerical inverse transformation method, see Proposition 1. Next, ρ_i and η_i are transformed to the original space by inverting the radial transformation and the location-scale transformation. In Table 1, we summarize the steps of one iteration of RMHS.

[3] Henceforth, the index i (in x_i^*, η_i^* ...) does not indicate the ith element of the corresponding vector, but indicates the number of the draw in a sequence of successive draws.

Step 4 of an RMHS iteration requires the acceptance probability $\alpha(\eta_{i-1}, \eta_i^*)$, and step 5 requires the distribution of the distance ρ conditional on the direction η_i. They are given in the next proposition.

Proposition 1. *For all elliptically contoured candidate distributions with mean μ and positive definite covariance matrix Σ, the acceptance probability of step 4 of the RMHS algorithm, summarized in Table 1, depends only on the generated direction and not on the functional form of the candidate density. The value of $\alpha(\eta_{i-1}, \eta_i^*)$ is given by*

$$\alpha(\eta_{i-1}, \eta_i^*) = \min\left\{\frac{I(\eta_i^*)}{I(\eta_{i-1})}, 1\right\}, \tag{21}$$

where

$$I(\eta) = \int_{-\infty}^{\infty} \kappa(\rho|\eta)\, d\rho, \tag{22}$$

and where $\kappa(\rho|\eta)$ is a kernel of the conditional density $p(\rho|\eta)$ of step 5, defined by

$$p(\rho|\eta) \propto \kappa(\rho|\eta) = p(x(\rho, \eta|\mu, \Sigma))|J_y(\rho)|. \tag{23}$$

Proof. See Appendix A. □

Remark 1. A noteworthy property is that the acceptance probability does not depend on the functional form of the candidate density under the condition that this candidate density is of the form $c_m(\det(\Sigma))^{-1/2} g((x-\mu)'\Sigma^{-1}(x-\mu))$, i.e. an elliptically-contoured density. However, the acceptance probability depends on the generated direction η and thus on the location and scaling matrix of the candidate density.

Remark 2. In order to obtain the acceptance probability $\alpha(\eta_{i-1}, \eta_i^*)$, the integral $I(\eta)$ defined by (22) can be computed by a deterministic integration rule. Since the density of ρ conditional on η is proportional to the integrand of $I(\eta)$, evaluations of the integrand, gathered during the deterministic integration phase, can be used in order to construct a grid for $p(\rho|\eta)$. Using the numerical inverse transformation method, sampling the distance ρ conditional on the direction η, that is, step 5 of an RMHS iteration, is straightforward.

Remark 3. The integral $I(\eta)$ has infinite integration bounds. However, in practice we use finite integration bounds for its numerical evaluation. Of course it is important to take these bounds such that practically all density mass of ρ given η is included, but for the sake of efficiency it is also desirable that the integration interval is as small as possible. In order to obtain bounds for the distance ρ we impose minimum and maximum values for each element of x in the original space. It is often possible to find sensible bounds by either theory and/or common sense, and in the applications below we will explicitly report the considered bounds. More generally, by imposing bounds on the values of x we impose linear inequality conditions on the original

parameter space. These restrictions are of the form $c_j'x \leqslant b_j$, where j indicates the number of the restriction. The conditions imposed on the original space translate to bounds $\rho_{\min}$ and $\rho_{\max}$ for ρ through the relationships $\rho_{\min} = \max_j\{\rho_j : \rho_j < 0\}$ and $\rho_{\max} = \min_j\{\rho_j : \rho_j > 0\}$, where $\rho_j = (b_j - c_j'\mu)/c_j{}'(\tilde{x} - \mu)$ with $\tilde{x} = x(\rho = 1, \eta|\mu, \Sigma)$. As additional linear restrictions might reduce the integration interval for $I(\eta)$, making the evaluation of $I(\eta)$ more efficient, they do not put a burden on the algorithm, but they might result in an efficiency gain.

2.2.2. Convergence

RMHS is a combination of a Metropolis-Hastings sampler for the directions η and the inverse transformation method for generating the distance ρ. The MH step on the directions introduces dependence on past drawings. The following theorem provides a sufficient condition for convergence of RMHS.

Proposition 2. *If Σ is non-singular and if for every $\eta \in \{\eta \in \mathbb{R}^{m-1} : \eta'\eta < 1\}$ in the support of the target density $p(\eta)$ the candidate density $q(\eta)$ is non-null and continuous, then the sampled RMHS chain converges in distribution to the target distribution.*

Proof. RMHS is a combination of a Metropolis-Hastings sampler for the directions η and direct generation of the distance ρ. Hence, the transition kernel of RMHS is the transition kernel of the MH step, and we can rely on known convergence results for the MH algorithm, see e.g. Smith and Roberts (1993) and the references cited there. In particular, if the covariance matrix Σ is non-singular, these convergence results are preserved after applying the location-scale transformation. Moreover, they are also preserved after applying the radial transformation given that this transformation does not induce singularities, which is the case if $\eta \neq \pm 1$ and $\rho \neq 0$. As these singularities have Lebesgue measure zero, the radial transformation does not affect convergence properties. So, under the conditions stated in Proposition 2, the sampled RMHS chain converges in distribution to the target distribution. Note that in Appendix A it is shown that for an elliptically contoured candidate, the density $q(\eta)$ is given by $q(\eta) \propto J_y(\eta) = (1 - \eta'\eta)^{-1/2}$. As $q(\eta)$ is clearly positive and continuous on $\{\eta \in R^{m-1} : \eta'\eta < 1\}$, the sufficient condition in Proposition 2 is satisfied. □

In practice, convergence should always be checked using diagnostics. One can monitor the sequential values of posterior moments, the acceptance rates, and the (tail) distribution and variance of the importance function $I(\eta)$, see (22).

Remark 4. In practice we reduce the computational effort by generating several drawings of ρ for each drawing of η, i.e. we capitalize on the construction of the grid for $p(\rho|\eta)$ (see Remark 2). We emphasize that the computed integrals (MC estimators) still converge to the theoretical integrals. The main point is that although the generated drawings of y and x are dependent, the computed integrals are consistent estimates of the theoretical values of the integrals that one is interested in, see Geweke (1999, p. 44) and the references cited there.

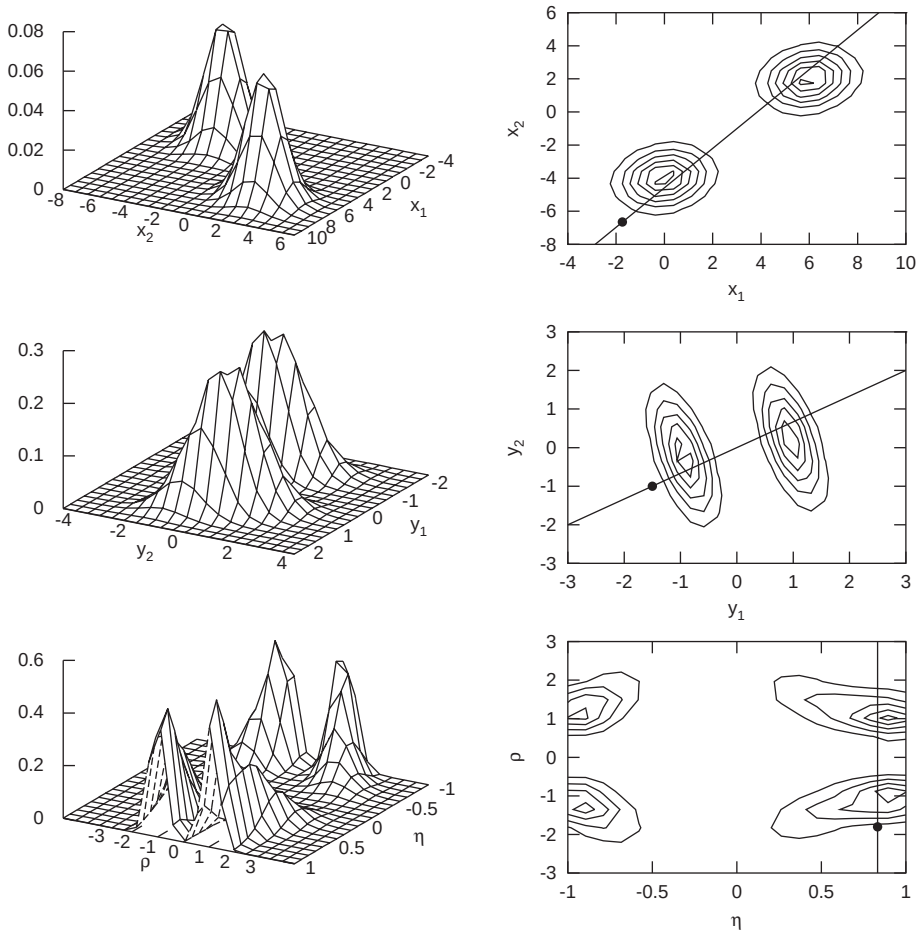

Fig. 3. Adaptive radial-based direction sampling: target density in original space (above), target density after location-scale transformation (middle) and target density after radial transformation (below).

2.2.3. *Illustration*

Fig. 3 illustrates RMHS for a bivariate bimodal target distribution. The upper two graphs display the target density in the original space. A point, representing a realization from the normal candidate distribution $N(\mu, \Sigma)$, is visible in the contour plot. If μ and Σ coincide with the mean and the covariance matrix of the target distribution, then the location-scale transformation leads to the target density that is depicted in the middle graphs. The gain of the location-scale transformation is clear: the density mass is better located around the origin in the sense that a line through the origin, defined by some direction η, 'hits the density mass' more easily. Since RMHS precisely considers such lines, the location-scale transformation may lead to a substantial improvement

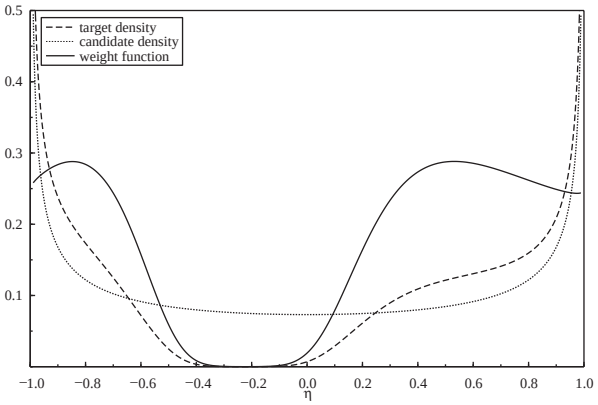

Fig. 4. The marginal target density $p(\eta)$, the marginal candidate density $q(\eta)$, and the weight function $p(\eta)/q(\eta)$ (up to scaling constants). The original target density is the same as in the previous figure.

for appropriate μ and Σ. The target density after applying the radial transformation is depicted in the bottom two graphs.

Seven steps are distinguished in an iteration of RMHS, see Table 1. The visualization of these steps in Fig. 3 is as follows. In step 1, the point in the upper contour plot is drawn from $\mathrm{N}(\mu, \Sigma)$. This point is transformed in step 2 to the point in the middle contour plot. Step 3 results in the point in the bottom contour plot. In step 4, the direction η, that is the horizontal position of the point in the bottom contour plot, is either accepted or rejected. In case η is accepted, step 5 consists of drawing one (or several) distance(s) ρ on the vertical line through the point. Step 6 can be represented by the transformation of points generated on the line in the bottom contour plot to points generated on the line in the middle contour plot. Similarly, step 7 results in points generated on the line in the upper contour plot. In sum, RMHS amounts to drawing from the exact target density on lines which are obtained through an MH acceptance-rejection mechanism. This mechanism takes into account the shape of the target density as we explain below.

Fig. 4 shows the marginal target density $p(\eta)$ that results from the bottom two graphs in Fig. 3. It looks quite ill behaved. However, similar to the relationship between standard independence chain MH and standard importance sampling, it is not directly the target density that matters for the MH step in RMHS. What matters is the ratio of the target density and the candidate density of η, that is $p(\eta)/q(\eta) \propto I(\eta)$ where $q(\eta)$ denotes the marginal candidate density and $I(\eta)$ has been defined in Proposition 1, see Appendix A for details. It is seen from Fig. 4 that, although the target density $p(\eta)$ is ill behaved, the weight function $p(\eta)/q(\eta)$ is quite well behaved in the sense that variation in its values is only moderate. We further note from the bottom two graphs in Fig. 3 that the transformed target density is bimodal (ill behaved) with respect to the distance ρ. However, this is not a problem, as a grid for $p(\rho|\eta)$ can always be constructed easily, see also Remark 2.

2.2.4. Adaptive radial-based Metropolis-Hastings sampling

For implementation of RMHS, the mean μ and the covariance matrix Σ of the normal candidate distribution have to be specified. Good enough initial approximations are usually the posterior mode and minus the inverse Hessian of the log posterior evaluated at the mode. Heuristically, convergence of RMHS should improve if μ and Σ are close to, rather than far from, the target mean and covariance matrix, respectively. Actually, if the target density is (approximately) normal, the acceptance probability should be (approximately) equal to 1 for appropriate μ and Σ. ARMHS considers a sequential adaptive approach. Given a generated sample $x_1, x_2, \ldots, x_n$ from a previous run of the algorithm, μ and Σ are replaced by the Monte Carlo estimates of the posterior mean and covariance matrix, which are given by

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i, \tag{24}$$

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} (x_i - \hat{\mu})(x_i - \hat{\mu})', \tag{25}$$

respectively. Using these estimates, one can proceed with a new sampling round. This process can be repeated any number of times. We note that information, coming from a 'wrong' sample, may have a misleading effect and may worsen convergence. Thus, convergence should be monitored by usual tools, see Van Dijk and Kloek (1980) and Oh and Berger (1992). However, since only the direction η, and not the distance ρ, is generated from the candidate distribution, the risk of collecting a 'wrong' sample is reduced. ARMHS should be quite robust, as the distance ρ conditional on the direction η immediately comes from the target distribution, that is, sampling on a given line mimics exactly the target density.

In order to monitor convergence over sampling rounds, we find the Mahalanobis distance particularly useful. It is defined as $\text{Mah}_j = (\hat{\mu}^{(j)} - \hat{\mu}^{(j-1)})'[\hat{\Sigma}^{(j)}]^{-1}(\hat{\mu}^{(j)} - \hat{\mu}^{(j-1)})$, where j indicates the sampling round. The Mahalanobis distance measures the extent to which the estimated posterior mean changes between successive sampling rounds, while accounting for parameter uncertainty and the underlying correlation structure. In the applications below we will use the Mahalanobis distance extensively.

2.3. Adaptive radial-based importance sampling

Radial-based importance sampling (RIS) replaces the MH step of RMHS for the direction η by an importance sampling step. So, step 4 of an RMHS iteration (see Table 1) changes. In RIS, every sampled direction η_i is kept, a distance ρ_i is sampled conditional on it, and the resulting radial coordinates are transformed to a draw x_i in the original space, which is weighted according to the appropriate importance weight $w(\eta_i)$:

$$w(\eta_i) = \frac{p(\eta_i)}{q(\eta_i)} \propto I(\eta_i), \tag{26}$$

where $I(\eta)$ is defined by (22), see Appendix A.

An interpretation of RIS is that one samples from the target distribution on lines with directions being derived from the elliptically contoured candidate distribution. Each line receives a weight, indicating the importance of the underlying direction. The weight of a line is carried over to any realization on that line. Alternatively, one may interpret RIS as just a special case of standard importance sampling. A realization x in the original space is a function of a realization (ρ,η) in the transformed space, see (18), implying that the importance weight of (ρ,η) is also the importance weight of x. Taken together, step 1 to step 5 of a RIS iteration can be regarded as yielding a realization (ρ_i,η_i) from a candidate distribution with density

$$q_{imp}(\rho,\eta) = q(\eta)\,p(\rho|\eta), \tag{27}$$

and providing the importance weight

$$w(\rho,\eta) = \frac{p(\rho,\eta)}{q_{imp}(\rho,\eta)} = \frac{p(\eta)\,p(\rho|\eta)}{q(\eta)\,p(\rho|\eta)} = \frac{p(\eta)}{q(\eta)}, \tag{28}$$

which coincides with (26). We note that the importance function q_{imp} takes into account the *shape* of the target density through $p(\rho|\eta)$. This is not the case in standard importance sampling with importance function $q(\eta)q(\rho|\eta)$ in the transformed space, which explains that the importance function of RIS should be more efficient than the importance function in standard IS with a normal or Student t candidate density. As RIS can be interpreted as a special case of importance sampling, convergence properties of RIS follow directly from those for the latter method. The distribution of the weights $w(\eta)$, in particular, a (bounded) variance of $w(\eta)$, are important diagnostics. For details, see Geweke (1989).

Similar to ARMHS, the parameters μ and Σ of the location-scale transformation can be updated by replacing them by their Monte Carlo estimates. These estimates are given by

$$\hat{\mu}_w = \frac{\sum_{i=1}^{n} w(\eta_i)\; x_i}{\sum_{i=1}^{n} w(\eta_i)}, \tag{29}$$

$$\hat{\Sigma}_w = \frac{\sum_{i=1}^{n} w(\eta_i)(x_i-\hat{\mu}_w)(x_i-\hat{\mu}_w)'}{\sum_{i=1}^{n} w(\eta_i)}, \tag{30}$$

where $x_1, x_2, \ldots, x_n$ is the collected sample, and $w(\eta_1), w(\eta_2), \ldots, w(\eta_n)$ are the corresponding importance weights. We will refer to this adaptive extension of RIS as adaptive RIS (ARIS).

2.4. On related methods and a classification

The two-dimensional polar transformation is the basis of the well known method of Box and Muller (1958) for generating normal random variables. Consider the angle θ formed by the y_1-axis and a line passing through the origin. In standard polar coordinates, a point is described by this angle θ and its Euclidean distance from the origin. In the Box-Muller method, an angle (therefore a direction η) is generated from

a uniform distribution on the interval $(0, 2\pi)$, and a squared distance is generated from an exponential distribution. See e.g. Rubinstein (1981, p. 86–87). In radial coordinates, the point is described by the radial coordinate η and the signed distance ρ, see (9) and (10) with $m = 2$. The ARDS method extends the Box-Muller algorithm by generating directions η using an MH or IS step, where the candidate density is compared with a target density. Given a generated candidate direction η, distances ρ are generated from a very accurate numerical approximation to the target distribution of the distances. Note that this exact distribution is model specific. If the normal candidate density in the original space is a good approximation to the target density in that space then the probability of acceptance in the MH step is close to one and the weight in the IS step is relatively constant. Non-normality can be evaluated using the weights computed in the one-dimensional integration step; see Hop and Van Dijk (1992) and Monahan and Genz (1997).

It is also of interest to compare ARDS with the class of adaptive direction sampling (ADS) algorithms, see Gilks et al. (1994). Two well known members of ADS are the hit and run algorithm of Schmeiser and Chen (1991) and the snooker algorithm of Gilks et al. (1994). In ADS, directions are sampled in the original parameter space. Only information on the shape of the target density is used. ARDS generalizes ADS in two ways. Firstly, use is made of an MH or IS step where candidate and target are compared to generate a direction. Secondly, one generates a distance ρ from a numerically very accurate approximation to the target distribution. This step is not always spelled out in ADS. Since ARMHS and ARIS are members of the MH and IS class of Monte Carlo methods, convergence properties of these methods are well established. This is not so transparent for the ADS methods.

We emphasize again that within ARDS one can make use of any candidate that belongs to the family of elliptical distributions. A good estimate of the location and scale is important for generation of efficient directions, that is, directions that define lines which cover the region where the target has substantial probability mass, see e.g. the line in Fig. 3. We note that Monahan and Genz (1997) use the terminology spherical radial integration in this context.

The ARDS class comprises several algorithms. One may distinguish between rejection sampling, importance sampling and Metropolis-Hastings sampling as Monte Carlo integration methods. So far, we have experimented with ARMHS and ARIS. However, one may also define a radial-based rejection sampling algorithm (ARRES): the sampled direction is accepted if $w(\eta) > cu$ (and rejected otherwise), where u is uniformly drawn in $(0, 1)$, and c is a constant such that the importance function envelopes the target function. For such an algorithm, formulas (24) and (25) apply to the accepted transformed draws. Consider next the case where generating random drawings of ρ is replaced by only evaluating the unidimensional integral. We name this case deterministic integration with respect to ρ. One can combine this deterministic integration with respect to ρ with rejection sampling, importance sampling or Metropolis-Hastings sampling with respect to η and evaluate posterior moments and densities. For the case of importance sampling this has been done in the so-called mixed integration method of Van Dijk et al. (1985), compare also Monahan and Genz (1997), who use the term spherical radial integration. Thus these methods are special cases of the ARDS class

where the step of generating random drawings of ρ is reduced to evaluating only a unidimensional integral. The limitation of deterministic integration with respect to ρ is that one has to compute a different unidimensional integral for each moment of the target distribution, see Hop and Van Dijk (1992). When one needs only to compute first and second-order posterior moments, one may use the mixed integration methods efficiently.

We emphasize that the ARDS sampling methods can be applied to target densities where the region of integration is bounded by several inequality conditions. Linear inequality conditions play an important role in several econometric and statistical applications.

3. Applications

In this section, a set of models are used to illustrate the versatility of radial-based algorithms. We compare ARMHS and ARIS to the (independence chain) Metropolis-Hastings algorithm and importance sampling. In all examples, MH and importance sampling use the Student t distribution with 5 degrees of freedom as the candidate. We start with an artificial example, which is the bimodal density from the previously considered illustration. Next, we analyze two empirical examples that are used in the literature for illustrative purposes. The first empirical example involves a regression model with scale contamination in order to investigate a study from Justel and Peña (1996) concerning the oxidation of nitric acid in a plant. Secondly, we consider a mixture model as in Frühwirth-Schnatter (2001) for the analysis of economic growth in the USA. We note that in these empirical examples the mixture process refers to the data space. However, it is noteworthy that such mixture processes may give rise to bimodality, extreme correlation and skewness in the parameter space. Of course, when the sample of data is very large and the model is regular such ill behaviour of the posterior density will be less likely. The issue is that a researcher does not know this before the posterior analysis of a model.

3.1. The bimodal density from the illustration

As an experiment we consider the bivariate bimodal density which was used in the illustration of RMHS (see Figs. 3 and 4). The distribution is given by

$$0.5\,\mathrm{N}\left(\begin{pmatrix}0\\-4\end{pmatrix},\begin{pmatrix}1&0\\0&1\end{pmatrix}\right)+0.5\,\mathrm{N}\left(\begin{pmatrix}6\\2\end{pmatrix},\begin{pmatrix}1&0\\0&1\end{pmatrix}\right).$$

We compute the first and second moments using RMHS, RIS, MH and importance sampling with three different candidate densities. The candidates differ in their location and scaling parameters. They are given by

$$\text{(a) } t_5\left(\begin{pmatrix}3\\-1\end{pmatrix},\begin{pmatrix}10&0\\0&10\end{pmatrix}\right),$$

$$\text{(b) } t_5\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}25 & 0\\0 & 25\end{pmatrix}\right),$$

$$\text{(c) } t_5\left(\begin{pmatrix}0\\-4\end{pmatrix},\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}\right).$$

The first candidate has the location and scaling parameters of the true target density. Only the correlation parameter is incorrect. The second candidate density has incorrect location but also a relatively large scale so that MH and importance sampling consider candidate drawings over a sufficiently large region. The third candidate density coincides with the first mode of the target density, so that the scale is too small. For each method, the moment estimates are based on 1000 drawings. In RMHS and MH, these drawings are preceded by a burn-in period of 100 drawings. In RMHS and RIS, we set the minimum and maximum bounds at respectively -10 and 10 for both components.

The estimated first and second moments and the true values for these moments are reported in Table 2. Furthermore, Fig. 5 displays the underlying RMHS and MH samples. The main findings are as follows. For the first two candidates (a) and (b), all four methods come up with good estimates for the scaling parameters and the correlation parameter, although the estimates from RMHS and RIS are slightly more accurate. However, the location parameters estimated using RMHS and RIS are much more accurate than those obtained from MH and importance sampling. The reason for this is illustrated in Fig. 5. The RMHS plots concerning the candidate densities (a) and (b) display much more points than the corresponding MH plots, indicating that the acceptance rates for the RMHS samples are much higher than the acceptance rates

Table 2
Sampling results for the bimodal density considered in the illustration of RMHS

		RMHS	RIS	MH	IS	True
(a)	mean(x_1)	2.83	2.96	2.70	3.35	3.00
	mean(x_2)	−1.16	−1.05	−1.37	−0.71	−1.00
	stdev(x_1)	3.15	3.16	2.98	3.09	3.16
	stdev(x_2)	3.15	3.17	3.21	3.28	3.16
	corr(x_1,x_2)	0.91	0.91	0.87	0.91	0.90
(b)	mean(x_1)	2.94	2.97	2.41	2.21	3.00
	mean(x_2)	−1.01	−1.01	−1.59	−1.59	−1.00
	stdev(x_1)	3.12	3.15	3.11	3.06	3.16
	stdev(x_2)	3.17	3.21	3.21	3.16	3.16
	corr(x_1,x_2)	0.89	0.91	0.93	0.91	0.90
(c)	mean(x_1)	2.90	3.19	0.01	0.02	3.00
	mean(x_2)	−1.14	−0.82	−4.02	−4.00	−1.00
	stdev(x_1)	3.11	3.11	1.04	1.02	3.16
	stdev(x_2)	3.15	3.18	1.04	1.01	3.16
	corr(x_1,x_2)	0.90	0.91	−0.03	0.00	0.90

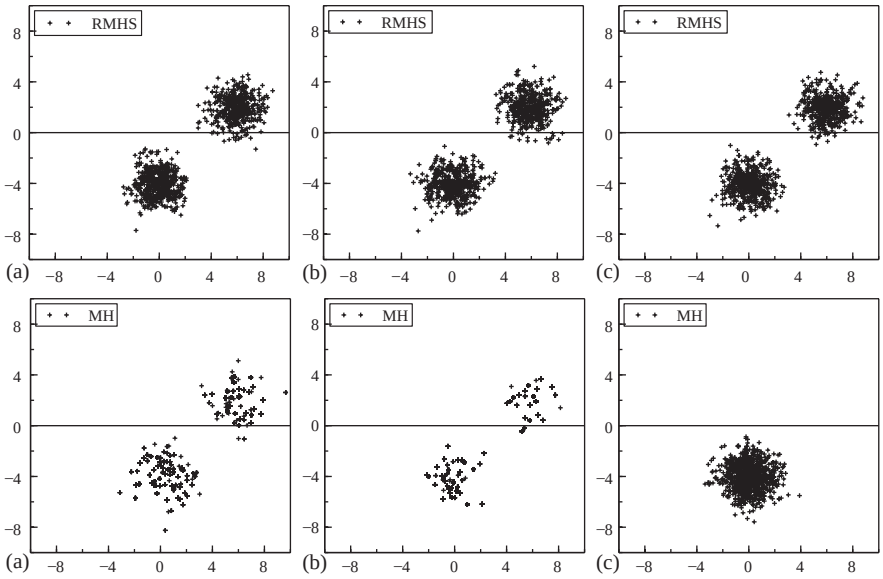

Fig. 5. Samples obtained for the bimodal density considered in the illustration of RMHS.

for the MH samples. This leads to a higher accuracy. Finally, it is seen for the third candidate (c) that MH and importance sampling fail to deliver a representative sample. They only find one mode and miss the other mode completely. This is in contrast to RMHS and RIS, which still succeed to come up with good moment estimates. In sum, this example demonstrates that, *given the same number of drawings*, ARDS algorithms should at least result in more accurate estimates than MH and importance sampling. Further, the example illustrates that ARDS algorithms should be more robust in the sense that the risk of overlooking density regions with non-trivial density mass (i.e. missing complete modes) is reduced. In this illustrative example, we conditioned on the number of drawings. However, as obtaining drawings from RMHS or RIS is more time consuming than obtaining drawings from the other two methods, we will make the computing times comparable in the subsequent empirical examples.

3.2. Scale contamination

Justel and Peña (1996) investigate a data set from Brownlee (1965, pp. 491–500) concerning the oxidation of ammonia to nitric acid in a plant. The data set incorporates 21 daily observations on four variables. The stack loss rate y, that is the proportion of ingoing ammonia to the plant that escapes unabsorbed, is related to the amount of air flow x_1 (representing the rate of operation of the plant), the temperature of the cooling water x_2, and the concentration of the circulating acid x_3. In several investigations it was found that several observations might be classified as outliers, and therefore care should be taken in the analysis to allow for this. In a regression setting, it is sufficient

Table 3
Sampling results for the scale contamination model

	Bounds ARDS		Initialization		ARMHS		ARIS		MH		IS		Large sample	
	min.	max.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
β_1	−10.00	10.00	0.00	2.00	0.81	0.19	0.81	0.19	0.81	0.20	0.81	0.20	0.81	0.19
β_2	−10.00	10.00	0.00	2.00	1.01	0.52	1.00	0.55	1.01	0.58	1.02	0.57	1.01	0.54
β_3	−10.00	10.00	0.00	2.00	−0.61	0.09	−0.61	0.10	−0.61	0.10	−0.61	0.10	−0.61	0.10
σ	0.00	10.00	5.00	2.00	2.84	1.32	2.77	1.38	3.06	1.38	3.09	1.39	2.82	1.36
κ	1.00	10.00	5.00	2.00	3.40	2.35	3.48	2.39	3.60	2.48	3.70	2.58	3.48	2.40
p	0.00	1.00	0.50	0.20	0.47	0.34	0.51	0.34	0.43	0.34	0.43	0.34	0.49	0.34
Drawings per iteration ($\eta \times \rho$)					5000 × 5		5000 × 5		250 000		250 000			
Number of iterations					5		4		4		4			
Average time per iteration (in s)					39.8		37.6		43.1		47.0			
Mahalanobis distance					0.02		0.00		0.01		0.01			
Acceptance rate (in %)					58.2				11.4					
5% most influential weights (in %)							19.6				62.8			

to allow for scale contamination, as in the model

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i,$$

$$\varepsilon_i \sim \mathrm{N}(0, \sigma_i^2),$$

$$\sigma_i = \begin{cases} \sigma & \text{with probability } 1-p, \\ \kappa\sigma & \text{with probability } p. \end{cases}$$

For identification of the two variance regimes, we impose that $\kappa > 1$. The prior for the parameter vector $(\beta_1, \beta_2, \beta_3, \sigma, \kappa, p)'$ is chosen proportional to $[(1-p)\sigma + p\kappa\sigma]^{-1}$ within the parameter bounds reported in Table 3. The prior density is zero outside these bounds.

Estimates for the mean and the covariance matrix of the six model parameters are computed by considering drawings from ARMHS, ARIS, MH and importance sampling. This is done in several sampling rounds. In our adaptive approach, additional sampling rounds are considered as long as the Mahalanobis distance is larger than 0.02. However, we allow for at most 8 rounds. In any round, ARMHS and ARIS draw 5000 directions and 5 distances per direction, resulting in a sample of size 25 000. In order to make the computing times comparable, the MH and importance sampling algorithms are allowed to collect a larger sample of size 250 000. The scale of the initial candidate distribution is taken sufficiently large, so that MH and importance sampling can initially cover the whole density mass.

The parameter estimates are reported in Table 3, together with the corresponding large sample values (computed from 250 000 ARMHS drawings). It is seen that all four methods trace back the response parameters β_1, β_2 and β_3 accurately. However, this is not the case for the remaining three parameters. In particular, MH and importance

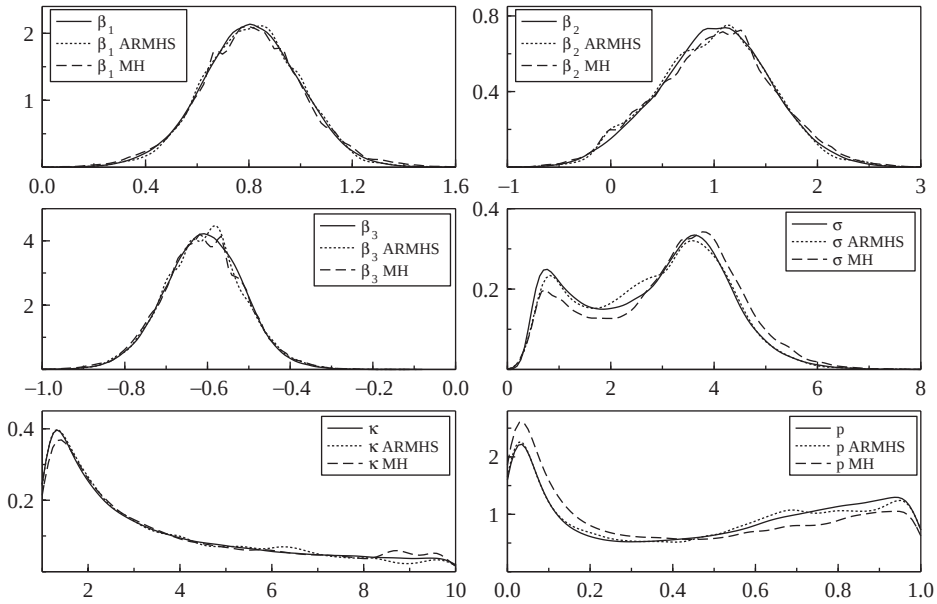

Fig. 6. Marginal densities for the scale contamination model.

sampling underestimate the left mode of the bimodal marginal density for σ, resulting in an overestimation of the mean of σ. This is also illustrated in Fig. 6 in which the large sample marginal densities (250 000 ARMHS drawings) and the estimated densities from 25 000 ARMHS drawings and from 250 000 MH drawings are shown. Furthermore, it is seen that MH and importance sampling underestimate the right mode of the bimodal marginal density for the mixture probability p, resulting in an underestimation of the mean of p. We note from Table 3 that the average times per sampling round are comparable for the four considered algorithms, and that they have all converged (in terms of our Mahalanobis distance definition) within 4 or 5 rounds.

The difference between the acceptance rates of ARMHS and MH is striking. The acceptance rate of the former is about five times as high as the acceptance rate of the latter. Further, the distribution of the importance weights for ARIS is much more uniform than the weight distribution for importance sampling. In the last round of the importance sampling procedure, the 5% most influential drawings have 63% of the total weight, whereas this is only about 20% in ARIS. Again, this demonstrates the accuracy of the ARDS algorithms.

3.3. A mixture model for the U.S. GNP growth rate

In models for the growth rate of the gross national product, great advances have been made by allowing for separate regimes in periods of recession and expansion. However, these models give rise to difficulties with respect to convergence of sampling methods

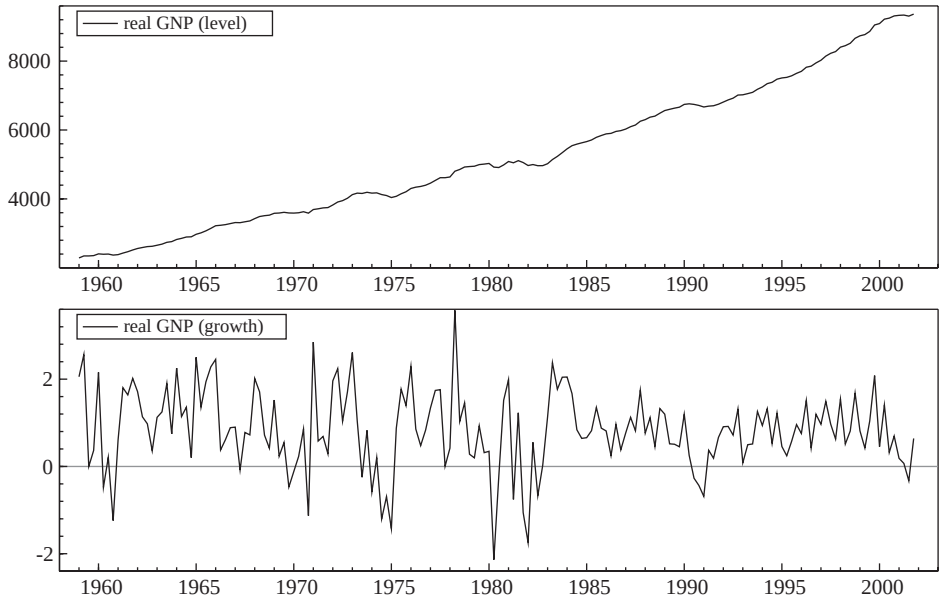

Fig. 7. Real GNP of the United States in billions of dollars (above), and its quarterly growth rate in % (below).

due to multiple modes. As an illustration of our algorithms we consider a mixture model with two AR(1) regimes for real GNP growth. This model is similar to the model considered in Frühwirth-Schnatter (2001) where another recent sampling method is discussed.[4] The model reads:

$$y_t = \begin{cases} \beta_{11} + \beta_{12} y_{t-1} + \varepsilon_t & \text{with probability } p, \\ \beta_{21} + \beta_{22} y_{t-1} + \varepsilon_t & \text{with probability } 1 - p, \end{cases}$$
$$\varepsilon_t \sim \mathrm{N}(0, \sigma^2), \tag{31}$$

where y_t denotes the quarterly growth rate. We investigate data concerning U.S. real GNP (source: Economagic). The data consist of observations from the first quarter of 1959 to the last quarter of 2001. Fig. 7 displays both the real GNP level and the quarterly growth rate (defined as 100 times the first difference of the logarithm). The priors for β_{11}, β_{12}, β_{21}, β_{22} and p are chosen uniform, and the prior for σ is taken proportional to $1/\sigma$. For identification, it is imposed that $\beta_{11} < \beta_{21}$. Again parameter bounds apply, see Table 4. The sampling setup is identical to the setup in the previously considered scale contamination example.

The parameter estimates are reported in Table 4, together with the corresponding large sample values (computed from 250 000 ARMHS drawings). Further, Fig. 8 shows the large sample marginal densities, and the densities estimated from 25 000 ARMHS

[4] Actually, Frühwirth-Schnatter (2001) also allows the variances to differ between regimes.

Table 4
Sampling results for the two-regime mixture model for U.S. real GNP

	Bounds		ARDS		Initialization		ARMHS		ARIS		MH		IS		Large sample	
	min.	max.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
β_{11}	−4.00	4.00	0.00	0.50	0.11	0.64	0.10	0.59	−0.14	0.88	0.01	0.72	0.07	0.70		
β_{12}	−1.00	1.00	0.00	0.50	0.45	0.24	0.45	0.25	0.42	0.28	0.40	0.28	0.41	0.27		
β_{21}	−4.00	4.00	1.00	0.50	1.32	0.74	1.27	0.78	1.22	0.83	1.28	0.85	1.30	0.79		
β_{22}	−1.00	1.00	0.00	0.50	−0.07	0.39	−0.02	0.38	0.05	0.39	0.01	0.40	−0.04	0.41		
σ	0.00	2.00	1.00	0.50	0.82	0.05	0.82	0.06	0.82	0.06	0.82	0.06	0.82	0.06		
p	0.00	1.00	0.50	0.20	0.59	0.38	0.53	0.38	0.48	0.39	0.52	0.39	0.55	0.38		

Drawings per iteration ($\eta \times \rho$)	5000 × 5	5000 × 5	250 000	250 000
Number of iterations	8	5	8	8
Average time per iteration (in s)	81.4	82.9	89.6	85.1
Mahalanobis distance	0.04	0.02	0.20	0.15
Acceptance rate (in %)	17.6		1.2	
5% most influential weights (in %)		57.9		99.7

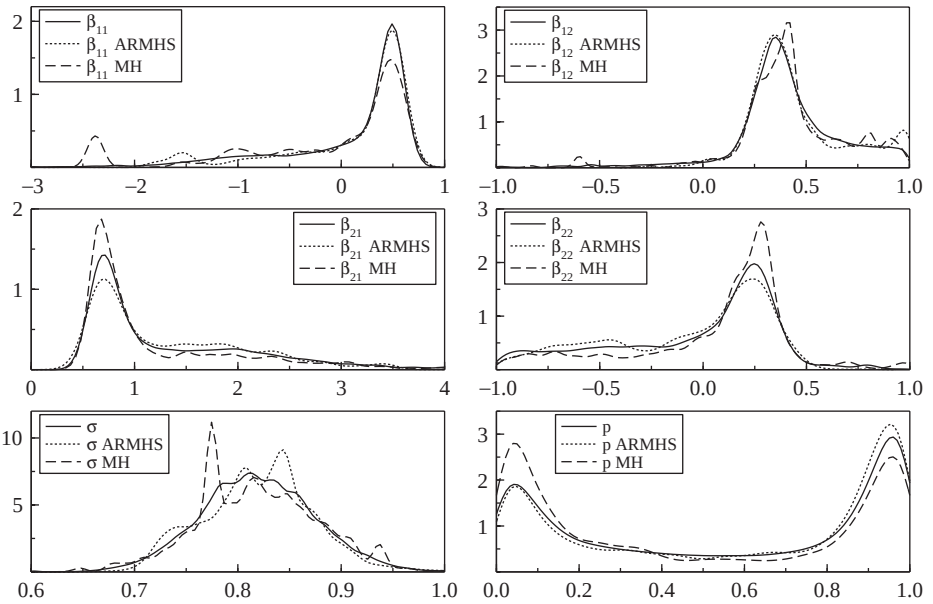

Fig. 8. Marginal densities for the two-regime mixture model for U.S. real GNP.

drawings and from 250 000 MH drawings. It is seen that the large sample densities (obtained from 250 000 ARMHS drawings) are smooth, while this is not always the case for the 'small sample' densities. However, we note that the 'small sample' MH densities have been constructed from the same number of drawings as the smooth large

sample ARMHS densities. In general, it is seen that the ARMHS, ARIS and importance sampling estimates are quite close to each other, but the MH estimates are sometimes quite different. In particular, this is illustrated in the density of the mixture probability p: MH overestimates the left mode and underestimates the right mode. The average times per sampling round are comparable for the four considered algorithms, but only ARIS 'converged' within 8 rounds, although the Mahalanobis distance for ARMHS is also quite small. The acceptance rates (ARMHS and MH) and the total weights of the 5% most influential drawings (ARIS and importance sampling) again provide support for the ARDS algorithms.

4. Conclusions

We have extended the Metropolis-Hastings and importance sampling methods by applying a radial transformation to the parameter space of the posterior (or target) density. Sampling does not take place in the m-dimensional parameter space directly, but in an $(m-1)$-dimensional subspace of directions. The final dimension, which corresponds to a distance measure, is sampled exactly from the target density (conditional on the directions), using the inverse transformation method. In this way the shape of the posterior density is taken into account perfectly along the sampled directions. For a given number of draws, this approach requires more functional evaluations of the posterior density than a traditional MH or IS algorithm. The usual type of tradeoff occurs: with a more sophisticated algorithm, one can hope to get 'correct' results with less draws than with a less sophisticated algorithm. It may also happen that a simple method cannot deliver reliable results. It would however be surprising when ARDS cannot deliver good results while the simpler, less computer intensive methods, can. This is confirmed by the (empirical) illustrations in Section 3. Moreover, a possible use of the ARDS algorithms is as a preliminary step to explore the posterior distribution and prepare a more sophisticated method. One may characterize our methods as 'black-box' algorithms. For details, see Chen et al. (2000).

We emphasize that there is no claim that ARDS algorithms are superior in theory to other kinds of algorithms (such a claim would make no sense). We believe that for any model/data combination, a sufficient research effort will usually allow to find a specific algorithm that performs better than ARDS or other algorithms. However, this is not necessarily guaranteed, and the specific algorithm may not be better even for a different data set (with the same model).

An interesting extension of this paper would be to embed ARMHS in a Gibbs algorithm, where a subset of the parameters can be directly simulated from their conditional distribution, while the remaining parameters cannot. In this framework, special care should be given to start sampling with sufficiently good initial guesses of the location and scale of the conditional distribution to be simulated. Examples where such an algorithm may be of great potential efficiency are in the Bayesian analysis of a linear simultaneous equation model, where the so-called simultaneity parameters induce a very non-elliptical shape of the posterior; a cointegration model, and a state space model as in Koop and Van Dijk (2000).

Acknowledgements

We thank Rodney Strachan, several referees, and participants of seminars at Cambridge University, CORE, Tinbergen Institute Rotterdam, ESEM, and STOCOM meetings for comments on an earlier version of this paper. We are, in particular, indebted to Michel Lubrano for very helpful comments which led to a substantial revision.

Support from HCM grant ERBCHRXCT 940514 of the European Commission is gratefully acknowledged. This paper presents research results of the Belgian Program on Interuniversity Poles of Attraction initiated by the Belgian State, Prime Minister's Office, Science Policy Programming. The scientific responsibility is assumed by the authors.

Appendix A. Proof of Proposition 1

First, given (18) and (19), the target density $p(x)$ in terms of ρ and η is given by

$$p(\rho,\eta) = p(x(\rho,\eta|\mu,\Sigma))|J_x(\rho,\eta)| = p(x(\rho,\eta|\mu,\Sigma))|J_y(\rho)|J_y(\eta)\det(\Sigma^{1/2}), \quad (A.1)$$

implying that

$$\begin{aligned} p(\eta) &= \int_{-\infty}^{\infty} p(\rho,\eta)\,d\rho = \int_{-\infty}^{\infty} p(x(\rho,\eta|\mu,\Sigma))|J_y(\rho)|\,J_y(\eta)\det(\Sigma^{1/2})\,d\rho \\ &\propto J_y(\eta)\int_{-\infty}^{\infty} p(x(\rho,\eta|\mu,\Sigma))|J_y(\rho)|\,d\rho \end{aligned} \quad (A.2)$$

(the last expression being a kernel of the marginal target density of η). Second, the elliptically contoured candidate density $q(x) = c_m(\det(\Sigma))^{-1/2}g((x-\mu)'\Sigma^{-1}(x-\mu))$ becomes the following function in terms of ρ and η:

$$\begin{aligned} q(\rho,\eta) &= q(x(y(\rho,\eta)|\mu,\Sigma))|J_x(\rho,\eta)| \\ &\propto g(\rho^2)|J_y(\rho)|J_y(\eta), \end{aligned} \quad (A.3)$$

so that

$$q(\eta) \propto J_y(\eta). \quad (A.4)$$

We note that (A.3) implies that ρ and η are independent. It follows from (A.2) and (A.4) that the acceptance probability $\alpha(\eta_{i-1},\eta_i^*)$, defined for an independence chain as

$$\alpha(\eta_{i-1},\eta_i^*) = \min\left\{\frac{\frac{p(\eta_i^*)}{q(\eta_i^*)}}{\frac{p(\eta_{i-1})}{q(\eta_{i-1})}},1\right\}, \quad (A.5)$$

simplifies to the expression in (21). Further, it follows from (A.1) that the density of ρ conditional on η, that is $p(\rho|\eta) \propto p(\rho,\eta)$, is given by (23).

References

Bauwens, L., Bos, C.S., Van Dijk, H.K., Van Oest, R.D., 2002. Adaptive polar sampling: a class of flexible and robust Monte Carlo integration methods. Working paper, Econometric Institute Erasmus University Rotterdam.

Box, G.E.P., Muller, M.E., 1958. A note on the generation of random normal deviates. Annals of Mathematical Statistics 29, 610–611.

Brownlee, K.A., 1965. Statistical Theory and Methodology in Science and Engineering, 2nd Edition. Wiley, New York.

Chen, M.-H., Shao, Q.-M., Ibrahim, J.G., 2000. Monte Carlo Methods in Bayesian Computation. Springer, New York.

Frühwirth-Schnatter, S., 2001. Markov chain Monte Carlo estimation of classical and dynamic switching models. Journal of the American Statistical Association 96, 194–209.

Geweke, J., 1989. Bayesian inference in econometric models using Monte Carlo integration. Econometrica 57, 1317–1339.

Geweke, J., 1999. Using simulation methods for Bayesian econometric models: inference, development, and communication. Econometric Reviews 18, 1–73.

Gilks, W.R., Roberts, G.O., 1996. Strategies for improving MCMC. In: Gilks, W.R., Richardson, S., Spiegelhalter, D.J. (Eds.), Markov Chain Monte Carlo in Practice, Chapman & Hall/CRC, New York, Boca Raton.

Gilks, W.R., Roberts, G.O., George, E.I., 1994. Adaptive direction sampling. The Statistician 43, 179–189.

Givens, G.H., Raftery, A.E., 1996. Local adaptive importance sampling for multivariate densities with strong nonlinear relationships. Journal of the American Statistical Association 91, 132–141.

Hobert, J.P., Casella, G., 1996. The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. Journal of the American Statistical Association 91, 1461–1473.

Hop, J.P., Van Dijk, H.K., 1992. SISAM and MIXIN: Two algorithms for the computation of posterior moments and densities using Monte Carlo integration. Computer Science in Economics and Management 5, 183–220.

Justel, A., Peña, D., 1996. Gibbs sampling will fail in outlier problems with strong masking. Journal of Computational & Graphical Statistics 5, 176–189.

Kleibergen, F.R., Van Dijk, H.K., 1994. On the shape of the likelihood/posterior in cointegration models. Econometric Theory 10, 514–551.

Kleibergen, F.R., Van Dijk, H.K., 1998. Bayesian simultaneous equations analysis using reduced rank structures. Econometric Theory 14, 701–743.

Koop, G., Van Dijk, H.K., 2000. Testing for integration using evolving trend and seasonal models: a Bayesian approach. Journal of Econometrics 97, 261–291.

Monahan, J., Genz, A., 1997. Spherical-radial integration rules for Bayesian computation. Journal of the American Statistical Association 92, 664–674.

Muirhead, R.J., 1982. Aspects of Multivariate Statistical Theory. Wiley, New York.

Oh, M.S., Berger, J.O., 1992. Adaptive importance sampling in Monte Carlo integration. Journal of Statistical Computation and Simulation 41, 143–168.

Rubinstein, R., 1981. Simulation and the Monte Carlo Method. Wiley, New York.

Schmeiser, B., Chen, M.-H., 1991. General hit-and-run Monte Carlo sampling for evaluating multidimensional integrals. Working paper, School of Industrial Engineering Purdue University.

Smith, A.F.M., Roberts, G.O., 1993. Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. Journal of the Royal Statistical Society, Series B 55, 3–23.

Van Dijk, H.K., Kloek, T., 1980. Further experience in Bayesian analysis using Monte Carlo integration. Journal of Econometrics 14, 307–328.

Van Dijk, H.K., Kloek, T., 1984. Experiments with some alternatives for simple importance sampling in Monte Carlo integration. In: Bernardo, J.M., Degroot, M., Lindley, D., Smith, A.F.M. (Eds.), Bayesian Statistics, Vol. 2. Amsterdam, North Holland.

Van Dijk, H.K., Kloek, T., Boender, C.G.E., 1985. Posterior moments computed by mixed integration. Journal of Econometrics 29, 3–18.