ADAPTIVE MARKOV CHAIN MONTE CARLO ALGORITHMS WITH GEOPHYSICAL APPLICATIONS

JOHANNA TAMMINEN

DEPARTMENT OF MATHEMATICS AND STATISTICS
FACULTY OF SCIENCE
UNIVERSITY OF HELSINKI
HELSINKI, FINLAND

ACADEMIC DISSERTATION in applied mathematics

To be presented, with the permission of the Faculty of Science of the University of Helsinki, for public critisism in Auditorium Exactum B123 (Gustaf Hällströmin katu 2b) on October 29th, 2004, at 10 a.m.

Abstract

The Markov chain Monte Carlo (MCMC) technique has during the last decade become a widely used approach to approximate complicated probability distributions. Such distributions are, for example, posterior distributions in inverse problems. The selection of a suitable MCMC algorithm is typically closely related to the choice of a suitable proposal distribution, which is known to be crucial for the effectiveness of this technique. The selection of the proposal distribution is not a trivial task in practice. The manual tuning of the proposal distribution is often laborous and time consuming. This is the case especially in high dimensional problems.

In this work we have developed adaptive MCMC algorithms to overcome the problem of selecting a sufficiently effective proposal distribution. The algorithms developed here are easily applicable, effective and, in some sense, 'generic' Metropolis-Hastings type MCMC algorithms which can be applied to a variety of different problems. They are also tested using different target distributions in various dimensions. In particular, in publication II we introduce the Adaptive Metropolis algorithm. It is by definition a non-Markovian algorithm, which is capable of using information of all the sampled points in tuning the proposal distribution. The algorithm, though non-Markovian, is shown to estimate correctly the underlying target distribution.

In addition to the development of adaptive MCMC algorithms the other objective of this dissertation is to apply the MCMC technique and especially the adaptive algorithms developed here to inverse problems of satellite remote sensing of the atmosphere. The MCMC technique has certain advantages over more traditional inverse techniques. These advantages include possibilities of solving nonlinear and non-Gaussian inverse problems. In addition, the MCMC technique allows flexibility in the definition of prior information and noise structure. In this work we have demonstrated these aspects by applying the MCMC technique to the inverse problems of the GOMOS (Global Ozone Monitoring by Occultation of Stars) instrument. The main purpose of the GOMOS instrument onboard the European Space Agency's Envisat satellite is to study the ozone depletion. We show that the MCMC technique is especially useful in validating the fast operational algorithms that are typically based on many assumptions and simplifications. These results can be used to improve the GOMOS measurements and their error estimates.

ILMATIETEEN LAITOS

Nimeke
Adaptiiviset Markovin ketju Monte Carlo -algoritmit ja niiden käyttö geofysikaalisissa sovellutuksissa

Tiivistelmä

Markovin ketju Monte Carlo (Markov chain Monte Carlo, MCMC) -menetelmän käyttö hankalien todennäköisyys-jakaumien approksimoimiseen on merkittävästi yleistynyt viimeisen kymmenen vuoden aikana. Tällaisia jakaumia ovat muun muassa käänteisongelmien posteriorijakaumat. MCMC-menetelmän tehokkuus on suuresti riippuvainen valitusta MCMC-algoritmista ja algoritmin käyttämästä ehdotusjakaumasta. Ehdotusjakauman optimointi kokei-lemalla on usein hyvin aikaa vievää ja työlästä mikä korostuu erityisesti korkeauloitteisissa ongelmissa.

Tässä työssä olemme kehittäneet adaptiivisia MCMC-algoritmeja, joiden tarkoituksena on helpottaa ehdotus-jakauman valintaa ja säätämistä. Kehittämämme algoritmit perustuvat Metropoliksen–Hastingsin algoritmiin, ja ne ovat helposti toteutettavia, tehokkaita ja soveltuvat moniin ongelmiin. Kaikki kehittämämme algoritmit on myös testattu moniulotteisissa testitapauksissa käyttäen useita eri kohdejakaumia. Julkaisussa II esittelemme erityisesti Adaptive Metropolis (AM) -algoritmin, joka käyttää hyväkseen kaikkia aikaisemmin arvottuja pisteitä ehdotus-jakauman säätöön. Vaikka AM-algoritmi ei ole markovinen, sen voidaan osoittaa approksimoivan kohdejakaumaa oikealla tavalla.

Adaptiivisten algoritmien kehittämisen lisäksi tämän työn toisena päämääränä on ollut soveltaa MCMC-menetelmiä ja erityisesti adaptiivisia algoritmeja ilmakehän kaukokartoitusmittausten tulkintaan. MCMC-menetelmän etuna muihin perinteisempiin tilastollisiin inversiomenetelmiin nähden on se, että sen avulla voidaan ratkaista suhteellisen tehokkaasti myös epälineaarisia ongelmia. Näitä joudutaan ratkaisemaan muun muassa kaukokartoitusmittausten tulkinnassa. Lisäksi MCMC-menetelmällä voidaan joustavasti ottaa huomioon a priori -tieto ja kohinarakenteet, jotka eivät noudata normaalijakaumaa. Tässä työssä olemme näyttäneet kuinka MCMC-menetelmää voidaan soveltaa GOMOS (Global Ozone Monitoring by Occultation of Stars) -mittalaitteen käänteisongelmiin. GOMOS on Euroopan avaruusjärjestön Envisat-satelliitissa toimiva mittalaite, jonka tärkein mittauskohde on otsonikato. Olemme näyttäneet myös kuinka MCMC- menetelmää voidaan käyttää hyväksi erityisesti validoitaessa GOMOSin nopeita operatiivisia algoritmeja, jotka perustuvat luonnollisesti useisiin oletuksiin ja yksinkertaistuksiin. Näitä tuloksia voidaan käyttää parantamaan GOMOSin mittaustuloksia ja niiden virhearvioita.

To my father who taught me what *infinity* means
and
to my mother who didn't think it was important

Preface

The work presented in this thesis has been carried out at the Finnish Meteorological Institute (FMI) and at the University of Helsinki, Department of Mathematics and Statistics, during 1998–2004. The main part of this work was done while I worked in FMI's Aeronomy group at the Department of Geophysical Research (GEO). In March 2004 GEO was re-named the Space Research Unit (AVA) and the Aeronomy group was moved to the Earth Observation Unit (KAU), where this work was revised.

I am first and foremost indebted to my advisors Prof. Heikki Haario from the Lappeenranta University of Technology and Prof. Eero Saksman from the University of Jyväskylä. Heikki's enthusiasm and wide knowledge in applied mathematics and experience of different applications have greatly impressed and helped me. Eero's profound understanding in mathematics and theoretical issues have been invaluably helpful to me during this work. It has been a privilege to be able to work with them.

I warmly thank Doc. Erkki Kyrölä, the head of the Aeronomy group first at GEO and now at KAU. His enthusiasm for atmospheric remote sensing, and unfailing support and patience have been important to me. He has continuously found time to discuss my work, carefully read through the manuscripts, and provided helpful criticism. I wish to express my sincere thanks also to Prof. Tuija Pulkkinen, the head of AVA (earlier GEO), who encouraged and helped me, especially, while writing the latter JGR-paper.

I also wish to thank my referees, Prof. Antti Penttinen and Prof. Erkki Somersalo for their positive criticism and valuable comments on this work.

I thank my co-authors Ph. Lic. Marko Laine and Prof. Markku Lehtinen for interesting ideas and enlighting discussions. I am grateful to Prof. Elja Arjas and Dr. Kari Auranen who helped me to get started with the MCMC technique. I thank also Harri Auvinen for enjoyable conversations on MCMC especially in high-dimensional problems.

The excellent working conditions at FMI have made this work possible. For this I wish to thank Prof. Risto Pellinen, the former head of GEO, Prof. Jarkko Koskinen, the head of KAU, Prof. Yrjö Viisanen, the head of the Research and Development Division and Professors Erkki Jatila and Petteri Taalas, the former and the present Director General of the Finnish Meteorological Institute, respectively

My thanks also go to Prof. Gilbert Leppelmeier for supporting me over the years and also for helping me with the English language. Moreover, I wish to thank my nearest colleagues in the Aeronomy group, Seppo Hassinen, Annika Seppälä, Viktoria Sofieva, and Pekka Verronen for pleasant working atmosphere. I am also grateful to my late colleague and friend Liisa Oikarinen, whose support

was indispensable during all the years we worked together.

I thank the European Space Agency for building and launching the Envisat satellite and especially the GOMOS instrument. Moreover, I thank all the people in Finland and abroad who have worked in the GOMOS project for years, and especially, Dr. Jean-Loup Bertaux at Service d'Aeronomie whose indefatigable enthusiasm for GOMOS made its realization possible.

In addition, a number of people at GEO/AVA deserve to be acknowledged for different reasons. Kirsti, Pekka J., Esa and Jouni P. for sharing their lunch breaks with me for more than 10 years; Johan, Pasi, Petri and Jouni R. for helping me with computers; Pete for being around; Ari-Matti for pressing me to complete the thesis; and Geodynamo for playing good music in our numerous parties.

I thank my family, relatives and all my friends for continuously reminding me of all the other important things in life than science. Finally, to my husband Reko and daughter Kristiina I would like to express my loving thanks for their unreserved support.

Helsinki, September 2004

Johanna Tamminen

Contents

LIST OF ORIGINAL PUBLICATIONS

I H. Haario, E. Saksman and J. Tamminen, (1999): Adaptive proposal distribution for random walk Metropolis algorithm. *Computational Statistics*, 14, 375–395.

II H. Haario, E. Saksman and J. Tamminen, (2001): An adaptive Metropolis algorithm. *Bernoulli*, 7(2), 223–242.

III H. Haario, E. Saksman and J. Tamminen, (2004): Componentwise adaptation for high dimensional MCMC. *Computational statistics*, Accepted.

IV J. Tamminen and E. Kyrölä, (2001): Bayesian solution for nonlinear and non-Gaussian inverse problems by Markov chain Monte Carlo method. *Journal of Geophysical Research*, 106(D13), 14,377–14,390.

V J. Tamminen, (2004): Nonlinear inverse algorithm validation with Markov chain Monte Carlo. *Journal of Geophysical Research*, Accepted.

VI H. Haario, M. Laine, M. Lehtinen, E. Saksman and J. Tamminen, (2004): MCMC methods for high dimensional inversion in remote sensing, with discussion. *Journal of the Royal Statistical Society, Series B*, 66, Part 3, 591–607.

1 Introduction

This thesis has two main objectives: development of adaptive Markov chain Monte Carlo (MCMC) algorithms and applying them in inverse problems of satellite remote sensing of the atmosphere. The motivation for developing adaptive MCMC algorithms originates in the practical problems that appeared while implementing the MCMC approach to the inverse problems of the GOMOS (Global Ozone Monitoring by Occultation of Stars) satellite instrument.

The adaptive MCMC algorithms, discussed in this thesis, focus on easily applicable, effective and, in some sense, 'generic' Metropolis–Hastings type MCMC algorithms. The intention has been to create algorithms that would work on a variety of problems with unknown posterior distributions. It is obvious that the posterior distributions of some problems are so complicated (e.g., multi-modal), that they require specifically tailored algorithms. Such problems are not considered in this thesis.

The MCMC technique has certain advantages over more traditional inverse techniques. These advantages include possibilities of solving nonlinear and non-Gaussian inverse problems. In addition, the MCMC technique allows flexibility in the definition of prior information and noise structure. In this work we have demonstrated these aspects by applying the MCMC technique to the inverse problems of the GOMOS satellite instrument. The implementation is possible only by using adaptive MCMC algorithms.

The thesis consists of 6 original publications which will be referred to by roman numerals (I–VI). The major contributions of the individual papers are as follows. PUBL. I introduces a practical, easy to implement random walk MCMC algorithm, Adaptive Proposal, which automatically searches for a proper proposal distribution for the MCMC algorithm and approximates the the underlying target distribution sufficiently well in many cases. In PUBL. II an adaptive MCMC algorithm, Adaptive Metropolis, is developed. The algorithm is the first fully non-Markovian MCMC technique for which the ergodicity is proven to hold. PUBL. III further develops a variant of the Adaptive Metropolis algorithm called the Single Component Adaptive Metropolis algorithm. This algorithm combines the ideas of single component sampling and the AM algorithm. In particular, high-dimensional problems are considered. In PUBL. IV the MCMC technique is introduced in the context of geophysical problems and, for the first time, applied to an atmospheric remote sensing problem. Results with simulated data are presented. PUBL. V introduces the methodology of using MCMC technique in the validation of operational data processing algorithms of atmospheric remote sensing instruments. It is shown that the MCMC technique can be used flexibly for validating and improving the operational algorithms. PUBL. VI considers two ways, both based on the adaptive MCMC techniques, for solving the posterior

distributions in a high-dimensional remote sensing problem.

This Summary of the thesis discusses in a general way the developed adaptive MCMC algorithms and the methodology of applying them to real inverse problems of satellite remote sensing. In addition, a short introduction to the GOMOS satellite instrument onboard the Envisat satellite is given.

2 Markov chain Monte Carlo technique for solving inverse problems

Indirect measurements are nowadays routinely used in natural sciences to study various physical and chemical phenomena which are difficult to observe using direct measurements. Examples of such measurements are, for instance, remote sensing measurements of the Earth and it's atmosphere and commonly used medical imaging techniques like X-ray and ultrasound measurements. In contrast to direct measurements the interpretation of the indirect measurements requires mathematical modeling and computational methods.

Let us denote by $y \in \mathbb{R}^m$ the measurements, by $x \in \mathbb{R}^d$ the unknown parameters that we are interested, and by f the relationship between these quantities. To interpret the indirect measurements we need to solve the inverse problem

$$y = f(x)$$

for x. Since the measurements include nearly always noise, it is natural to consider them and the unknown parameters as random variables. The Bayesian solution, i.e., the posterior distribution, is pointwise characterized by the posterior probability density function:

$$p(x \,|\, y) = \frac{p(x)p(y \,|\, x)}{\displaystyle\int p(x)p(y \,|\, x) \; dx}. \tag{2.1}$$

The posterior distribution combines the a priori information $p(x)$ and the measurement likelihood $p(y \,|\, x)$. To make inferences with respect to the posterior distribution we need to compute integrals of the form

$$\mathbb{E}[f(x)] = \int f(x)p(x \,|\, y) \; dx, \tag{2.2}$$

where f is some integrable function. The posterior distribution is typically characterized by computing the expectation, the probability regions of marginal posterior distributions, various quantiles, the covariance matrix and possibly higher moments which all require integration (2.2). In the general case this is a complicated task as no analytic solutions exist. Monte Carlo techniques, on the other hand, are based on approximating (2.2) by sampling $(X_1, \ldots, X_n)$ from the posterior distribution so that the expectation (2.2) with respect to the posterior distribution could be approximated by using finite sums:

$$\mathbb{E}[f(x)] \approx \frac{1}{n+1} \sum_{i=0}^{n} f(X_i).$$

In the traditional Monte Carlo sampling the states sampled are independent, but in the Markov chain Monte Carlo (MCMC) sampling they may be dependent

forming a Markov chain whose stationary distribution is the target posterior distribution $p(x \mid y)$. Some of the problems typically faced in the traditional Monte Carlo sampling are overcome in the MCMC technique; most importantly the posterior density (2.1) needs to be evaluated only up to a normalizing constant. In addition, the sampling can be efficient by concentrating on interesting areas since the samples are not independent.

Our main motivation in this work has been to apply the MCMC technique to solve inverse problems and to approximate posterior distributions. However, the MCMC technique and the adaptive algorithms (discussed in Chapter 3) can naturally be used to approximate also other distributions.

2.1 MCMC ALGORITHMS

The original idea of MCMC was introduced already 50 years ago in Metropolis, Rosenbluth, Rosenbluth, Teller and Teller [1953] where the algorithm was used in statistical physics to compute properties of substances consisting of interacting individual molecules. This algorithm has been extensively used in statistical physics [e.g., Hammersley and Handscomb, 1964] and appeared also in spatial statistics and statistical image analysis [e.g., Geman and Geman, 1984]. However, the utilization of MCMC algorithms for posterior inference was realized much later by Gelfand and Smith [1990]. Since then the MCMC technique has become a commonly used technique for approximating posterior distributions in a wide range of applications and several introductions to the technique have been published [Tierney, 1994; Gilks, Richardson and Spiegelhalter, 1996; Robert and Casella, 2000; Chen, Shao and Ibrahim, 2000]. The success and power of this technique are based on the simplicity of the basic MCMC algorithm. Another reason is due to the advances in computers: samples from the posterior distribution can now be computed in a reasonable time also for real problems.

Most of the MCMC algorithms are variants of the Metropolis–Hastings (MH) algorithm, which is based on the original Metropolis algorithm introduced in Metropolis, Rosenbluth, Rosenbluth, Teller and Teller [1953] and extended to cover also non-symmetric proposal distributions in Hastings [1970].

The MH algorithm is very simple: assuming that we have already sampled points $X_0, \ldots, X_{t-1}$ the algorithm proceeds in two steps. First a so-called candidate point Z is sampled from a proposal distribution q that may depend on the present point X_{t-1}. Next, the candidate point is either accepted or rejected using as the acceptance probability

$$\alpha(X_{t-1}, Z) = \begin{cases} \min\left(\dfrac{\pi(Z)q(Z, X_{t-1})}{\pi(X_{t-1})q(X_{t-1}, Z)}, 1\right) & \text{if } \pi(X_{t-1})q(X_{t-1}, Z) > 0, \\ 1 & \text{if } \pi(X_{t-1})q(X_{t-1}, Z) = 0. \end{cases} \quad (2.3)$$

where $q(x, z)$ denotes the probability of proposing z when at point x and $\pi(\cdot)$

stands for the density of the target distribution (i.e., posterior density $p(x\,|\,y)$ in inverse problems). In practice, the initial state X_0 is always chosen so that $\pi(X_0) > 0$.

The sampled chain that is used to approximate the posterior distribution has to be ergodic in the correct sense. The acceptance probability (2.3) of the MH algorithm is selected so that the chain is reversible. The reversibility ensures that the chain has the desired target distribution $\pi(x) = p(x\,|\,y)$ as the stationary distribution. The basis for the MCMC technique is given by the following theorem (using the formulation of [Nummelin, 2002]). Let us consider here such transition probability kernels P which consist of a singular part, i.e., the chain will stay put, and a continuous part, i.e., the chain will make a move (for a more exact definition see Nummelin [2002, Sec. 2.1]).

Theorem 1. (Law of large numbers for Markov chains) *Let $X_0, \ldots, X_n$ be a time-homogeneous Markov chain in the state space $E \subset \mathbb{R}^d$ with the transition probability P. Assume that the chain $X_0, \ldots, X_n$ satisfies the following two conditions:*

1) There exists a small set[1] $I \subset E$ such that for each initial state $x \in E$,

$$P^{n_x}(x, I) := P(X_{n_x} \in I \,|\, X_0 = x) > 0,$$

for some integer $n_x \geq 1$ depending on x.

2)   i) The chain $X_0, \ldots, X_n$ has a stationary probability density function $\pi(\cdot)$

*   ii) The support $S := \{x \in E \,:\, \pi(x) > 0\}$ of the stationary probability density function is closed in the sense that $P(x, S) = 1$ for all $x \in S$.*

Then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(X_i) = \int f(x)\pi(x)dx$$

for all π-integrable functions f and for all initial states $X_0 = x$ belonging to the support S of the stationary probability density function.

A self contained proof of this theorem can be found in Nummelin [2002]. A discussion on these issues is also given in Tierney [1994], which however, relays strongly on the results presented in Nummelin [1984]. An easy proof in the case f is bounded can be found in Tamminen [1999]. For MCMC algorithms the conditions in the theorem above are easily fulfilled.

Different choices for the proposal distribution q give rise to different sampling algorithms. A commonly used MH technique, called the random walk MH algorithm, refers to the case where the proposal distribution depends on the distance between the current point and the proposed point ($q(x, z) = g(x - z)$).

---

[1] A set $I \subset E$ with volume $|I| > 0$ is called small, if there exists a subset $J \subset E$ with volume $|J| > 0$ and a positive constant $\beta > 0$ such that $p(x, y) \geq \beta$ whenever $x \in I, y \in J$. Here $p(x, y)$ denotes the probability of moving from x to y.

A special case of this algorithm where the proposal distribution is symmetric ($g(x) = g(-x)$) leads to an algorithm that was originally proposed by Metropolis, Rosenbluth, Rosenbluth, Teller and Teller [1953]. This popular Metropolis algorithm involves only comparisons of the target function values at the present point and at the candidate point and it is therefore quite attractive in practice. A classical and widely used symmetric proposal is a Gaussian distribution centered at the current point. The algorithms developed in this work are also modifications of this traditional Metropolis algorithm with Gaussian proposals. Another class of widely used approaches are based on independence sampling. Here the proposal distribution (typically an approximation of the target distribution) does not depend on the current point ($q(x, z) = q(z)$). This type of algorithm is not discussed further in this work.

The sampling in MH algorithms may take place directly in a d-dimensional space or stepwise in a lower dimensional space, e.g., coordinate by coordinate as in the original Metropolis algorithm. The latter approach is nowadays known as the single component MH algorithm. The Gibbs sampling algorithm [Geman and Geman, 1984] can also be considered as a special case of the single component MH algorithm where the proposal distributions equal with the full conditional distributions. In this work both single component and multidimensional MH approaches are considered.

2.2 Comparing the performance of MCMC algorithms

The performance of a MCMC chain is often characterized by the speed of convergence and the efficiency of the chain [Besag and Green, 1993]. Roughly speaking, the speed of convergence can be understood as a measure of how quickly the algorithm converges to the target distribution, and the efficiency as the capability of the chain to explore the whole target distribution. Both of these can be addressed in terms of the spectrum of the Markov transition kernel and require computation of the eigenvalues of the transition kernel. In practice some approximations are used instead. One of the measures used for efficiency is the integrated autocorrelation value [Sokal, 1989]. It can be applied to study the efficiency of the one-dimensional projections of the chains.

In this work we have empirically tested the performance of different algorithms by comparing their capabilities to approximate certain known, linear and nonlinear, target distributions (Publ. I). The testing procedure is straightforward, but contains some novel features. The approach was motivated by the need to apply MCMC to real multidimensional problems with similar target distributions. A somewhat similar approach has later been used also by Warnes [2001].

The testing procedure applied to targets in varying dimensions is as follows: as target distributions we have used uncorrelated and correlated Gaussian dis-

tributions and twisted, 'banana-shaped', Gaussian distributions; see Fig. 2.1 for examples in 2-d. They have been selected so that analytical expressions could be used to compute different probability regions of the target distribution. In addition, they represent reasonably well typical shapes of posterior distributions in many inverse problems. Multi-modal distributions have not been considered in this work. Each test is repeated 10–100 times with varying starting points close to the mode of the target distribution. Finally, statistical analysis is performed to compute different performance criteria for the algorithm. As such criteria we used, e.g., the mean distance of the expectation values from the true value and the mean error of the percentages of the sampled points that are located inside some pre-defined probability region. The first criteria characterizes how well the expectation can be approximated and the second how well the posterior distribution is covered by the sampled points.

In the comparisons we have used essentially the same number of target function evaluations for each of the algorithms compared. This decision is based on the fact that in real life problems, nearly always the most time consuming part in the MCMC sampling is the evaluation of the target function $\pi(\cdot)$. Therefore, algorithms that approximate the target distribution more accurately using a given number of function evaluations can, roughly speaking, be considered as more efficient compared to the others.

2.3 NEED FOR ADAPTATION

Despite the simplicity of the basic MH algorithm, the implementation of the MCMC technique is not a straightforward procedure. In real problems, the performance of the theoretically ergodic MH algorithm may be far from acceptable, since reasonable results are needed in a finite time. Generally speaking, the performance of the MCMC technique depends on two things: the target distribution and the selected MCMC algorithm. Improvements in the efficiency can be achieved, e.g., by reparameterizing the target distribution, but this type of changes requires that the target distribution is known beforehand [Gilks and Roberts, 1996]. More practical improvements are thus related to the choice of the MCMC algorithm. In the context of MH type algorithms this relates to optimizing the size and the shape of the proposal distribution q.

It is well known that a good proposal distribution is crucial for the effectiveness of the MH sampling [e.g., Gelman, Roberts and Gilks, 1996]. A poor proposal distribution might result in a chain that does not represent the target distribution well even if run for a relatively long time. This is also demonstrated in Fig. 2.2 where the proposal distribution is either too small (top panel), too large (middle panel) or nearly optimal (bottom panel). The selection of the proposal distribution is typically done by trial and error using pre-runs as suggested
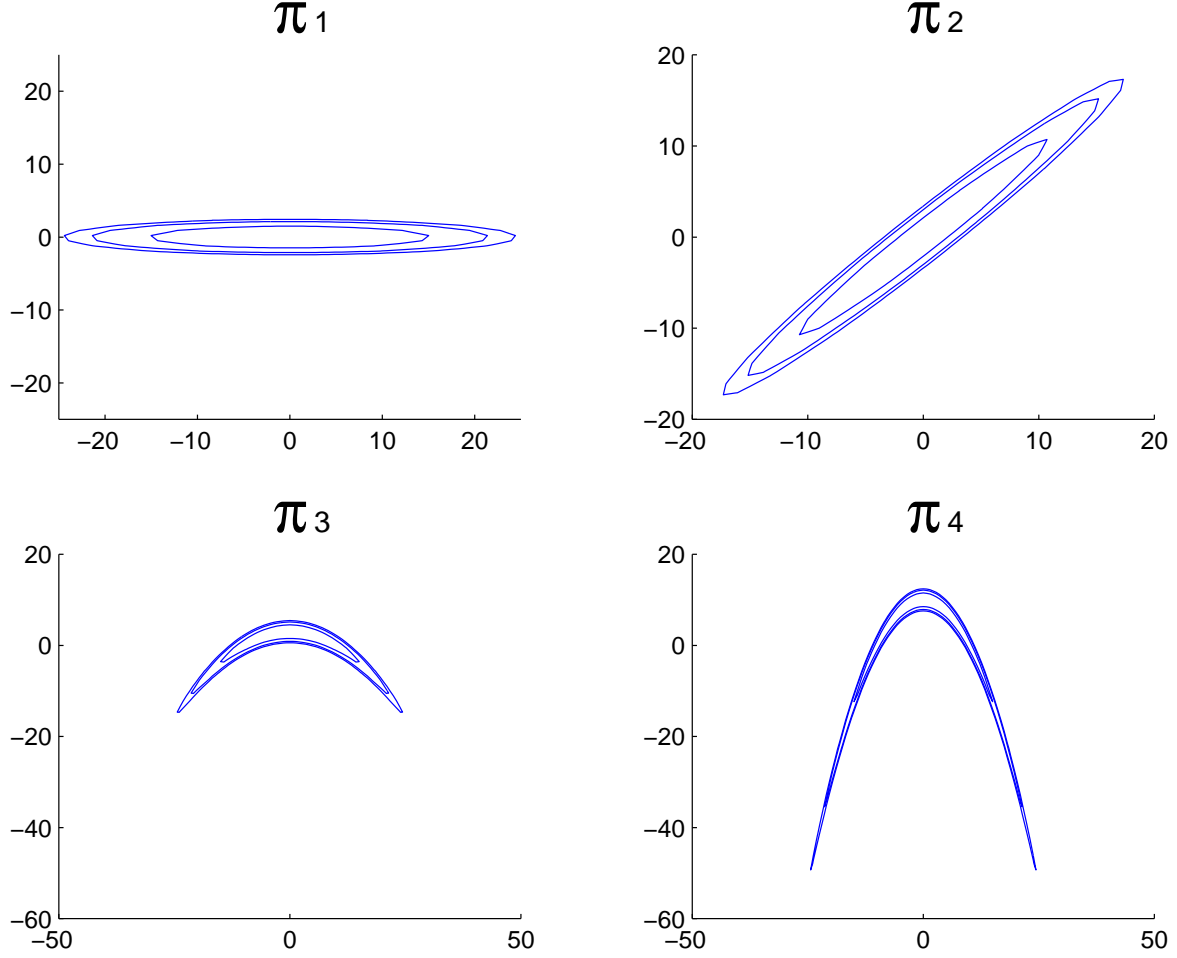
FIGURE 2.1. Target distributions used in the tests.

by Gelfand and Sahu [1994]. It is common to monitor the acceptance ratio and tune the proposal distribution to obtain some desired (ad-hoc) acceptance ratio, typically around 20-70 %. This manual tuning of the size and the shape of the (multidimensional) proposal distribution is a laborious and time consuming task. When the parameters are, for example, of different orders of magnitude and correlated, the tuning of the proposal distribution becomes complicated if based on simply monitoring the acceptance ratio. In high-dimensional problems this might even become impossible in practice. Therefore, automatic techniques for finding good proposal distributions are needed to make the MH algorithms applicable in practice. The manual tuning of the proposal distribution also turned out to be the bottleneck of implementing the MH algorithm to GOMOS inversion, as discussed more in Sec. 4.4.
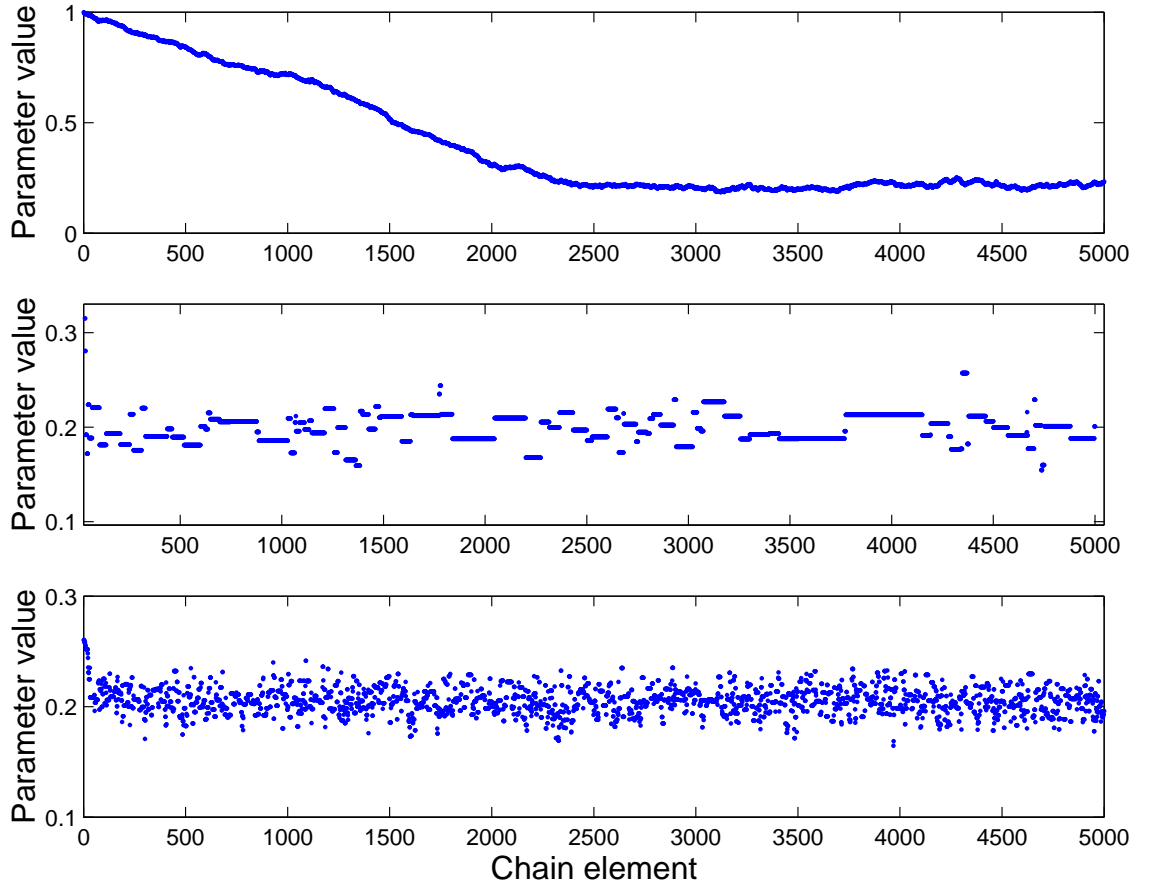
FIGURE 2.2. Examples of sampled points using MH algorithm with varying proposal distributions. Top panel: a too small proposal distribution results in accepting almost all points. Middle panel: a too large proposal distribution results in rejecting a large part of the proposal points. Bottom panel: a reasonable proposal: about 35% of the points are accepted in this example. The starting point was the same for all the algorithms (not shown in the two lowest panels).

3 ADAPTIVE MCMC

3.1 ADAPTIVE ALGORITHMS

Automatic techniques that use information collected during the MCMC sampling to improve the performance are called adaptive MCMC algorithms. During the last 10 years many adaptive MCMC algorithms have been proposed to optimize the performance of the standard MH algorithm [for further information, see e.g., PUBL. I–III; Gilks, Roberts and George, 1994; Gilks, Roberts and Sahu, 1998; Holden, 1998; Tierney and Mira, 1999; Warnes, 2001; Chauveau and Vandekerkhove, 2002; Gåsemyr, 2003; Sahu and Zhigljavsky, 2003; Andrieu and Robert, 2001; Atchade and Rosenthal, 2003; Erland, 2003, and the references therein]. The critical point in adaptive MCMC algorithms is that the adaptation may disturb the Markovian property so that the ergodicity of the algorithm is not guaranteed by the standard ergodicity theory of MCMC. Our aim here is not to make an extensive overview of adaptive MCMC techniques but rather to mention shortly the most relevant adaptive MCMC algorithms with respect to the algorithms developed in PUBL. I–III.

3.2 THE ADAPTIVE PROPOSAL ALGORITHM

A natural way of improving the proposal distribution is to use pre-runs and tune the proposal distribution based on the experience of the pre-runs as suggested in Gelfand and Sahu [1994]. Here the adaptation takes place only during the burn-in phase and after the tuning the proposal distribution is fixed. Since the adaptation is not continued after the burn-in phase the convergence is ensured by the basic theory. This simple approach has been used in many practical applications. The Adaptive Proposal (AP) algorithm introduced in PUBL. I can be considered to belong to this category, although it can also be thought as an approximately correct algorithm whose exactness is sufficient for many practical purposes.

The AP algorithm resembles the standard random walk Metropolis algorithm with the exception that the Gaussian proposal distribution q_t depends on time:

$$q_t(X_{t-1}, \cdot) = \mathbf{N}(X_{t-1}, s_d R_t(h))$$

where $R_t(h)$ corresponds to the empirical covariance matrix of h last points

$$R_t(h) = \text{cov}(X_{t-h}, \ldots, X_{t-1}).$$

The scaling $s_d = 2.4^2/d$ is chosen so that it is optimal in the case of a Gaussian target and a Gaussian proposal [Gelman, Roberts and Gilks, 1996]. The acceptance probability used in AP equals the Metropolis acceptance probability.

The AP algorithm is simple and easy to implement. The multivariate Normal proposal distribution takes naturally into account the possible correlations between the parameters. The additional computing time in the AP algorithm is rather small in low-dimensional problems. Numerous tests in Publ. I (see also [Haario, Saksman and Tamminen, 1998]) show that the AP algorithm can be used to approximate reasonably well behaving, low-dimensional, posterior distributions in many cases.

However, when the adaptation in the AP algorithm is continued after the burn-in period the correct ergodicity is not guaranteed. The stationary distribution of AP may actually be different from the target distribution. For many practical examples the difference is perhaps negligible, but for some special targets the difference is crucial, as demonstrated in Publ. I. Using the AP algorithm as an effective burn-in for ergodic MCMC algorithms may also be problematic. The adaptation during the burn-in phase may work well in some cases, but it is not guaranteed that a proper proposal distribution is found.

A similar idea of updating the covariance matrix of a Gaussian proposal distribution during the burn-in phase was also independently used by Hanson and Cunningham [1998]. Their adaptation, however, was based on applying a different numerical approach.

3.3 Continuously adaptive algorithms

In addition to the quasi-adaptive methods, like AP, fairly many adaptive algorithms have been proposed where the adaptation is continued also after the burn-in period. Techniques that rely on the standard theory of MCMC algorithms use only partly the history for adaptation. Many of these techniques are based on using multiple chains [e.g., Gilks, Roberts and George, 1994; Chauveau and Vandekerkhove, 2002; Warnes, 2001]. The practicality of these techniques (especially in high dimension) may be limited because of the memory requirements of multiple chains. For example, the technique by [Chauveau and Vandekerkhove, 2002] relays on running an increasing number of chains. Algorithms based on delaying rejection [Tierney and Mira, 1999] can be understood as locally adaptive methods. However, in the basic version of this method multiple (fixed) proposals are used rather than truly adaptive techniques.

Continuously adaptive MCMC algorithms (introduced so far) that use the whole history and only single chain for adaptation are either based on regeneration or slowing down the adaptation along the sampling (see also Erland [2003]). The latter type of adaptation is referred to as adaptation with diminishing effect in Erland [2003] and we will here also employ this terminology.

The regeneration idea is proposed by Mykland, Tierney and Yu [1995] and Gilks, Roberts and Sahu [1998]. The ergodicity is preserved by updating the pro-

posal distribution only when entering to a regeneration set. In real, multidimensional, problems the regeneration is rather complicated to ensure and therefore the practicality of this technique is restricted. The Self-regenerative algorithm with adaptation [Sahu and Zhigljavsky, 2003] resembles independence sampling with a proposal distribution that is a mixture of distributions. The adaptation takes place at so called trouble points and updates the proposal distribution by adding a new component to the mixture. The proposal distribution consists thus of an increasing number of distributions. In high-dimensional problems, again, the practicality of this algorithm might therefore be limited.

The adaptive MCMC algorithms, Adaptive Metropolis algorithm [PUBL. II] and the Single component adaptive Metropolis algorithm [PUBL. III], discussed in this work are based on using adaptive techniques with diminishing effect. This idea of adaptation was introduced in PUBL. II and generalized later, especially, in Andrieu and Robert [2001] (see also Atchade and Rosenthal [2003]).

3.4 THE ADAPTIVE METROPOLIS ALGORITHM

The Adaptive Metropolis (AM) algorithm is a fully non-Markovian MCMC algorithm in the sense that each step is a non-Markovian and the algorithm is able to use the whole history for adaptation [PUBL. II]. It is based on the same idea of updating the covariance matrix of the random walk Metropolis algorithm as in AP. In AM the covariance matrix is updated by using information of the whole history (or at least suitably increasing part) of the already sampled points. In AM the proposal distribution is

$$q_t(X_{t-1}, \cdot) = \mathbf{N}(X_{t-1}, C_t)$$

where C_t is defined as

$$C_t = \begin{cases} C_0 & \text{if } t \le t_0 \\ s_d \text{cov}(X_0, \dots, X_{t-1}) + s_d \varepsilon I_d & \text{if } t > t_0 \end{cases}$$

The scaling factor $s_d = 2.4^2/d$ corresponds to the same scaling as in the AP algorithm and I_d is a d-dimensional identity matrix. The role of the scaling ε is to prevent the covariance C_t becoming a singular matrix and it is chosen to be small (compared to the size of the support of π). The time t_0 reflects our trust on the initial covariance matrix C_0.

The AM algorithm is clearly non-Markovian and the ergodicity proof of standard MCMC can not be applied. However, the AM algorithm is ergodic and we have the following theorem.

Theorem 2. [PUBL. II] *Assume that the target density π has bounded support and is bounded from above. Then the AM chain simulates properly the target*

distribution: for any bounded and measurable function f *it holds almost surely that*

$$\lim_{n \to \infty} \frac{1}{n+1}(f(X_0) + f(X_1) + \ldots + f(X_n)) = \int f(x)\pi(x)dx.$$

Heuristically, the ergodicity follows from the fact that the changing of the proposal due to adaptation slows down in the course of sampling. This 'freezing' of the chain can be seen from the recursive formula for updating the covariance:

$$C_{t+1} = \frac{t-1}{t}C_t + \frac{s_d}{t}\left(t\overline{X}_{t-1}\overline{X}_{t-1}^T - (t+1)\overline{X}_t\overline{X}_t^T + X_tX_t^T + \varepsilon I_d\right),$$

where $\overline{X}_t$ stands for the empirical mean:

$$\overline{X}_t = \frac{1}{t+1}\sum_{i=0}^{t} X_i.$$

The proof of the theorem needs basically two things:

(i) The distribution of X_n approaches π as $n \to \infty$.

(ii) The dependence of the chain on fixed size time intervals of the past decreases along time.

To understand why (i) and (ii) are true, consider the time interval $I := (n, n+1, \ldots, n+j)$ where $n \gg j$. Along this interval the covariance C_n stays almost constant and the chain is approximately Markovian for $n \leq t \leq n+j$. Let X'_t, $t \geq n$ be the approximative chain obtained by setting $C'_t = C_n$ for $t \geq n$. Thus, we consider the chain

$$X_0, \ldots X_{n-1}, X_n, X'_{n+1}, \ldots, X'_{n+j}, \ldots \ldots \qquad j \ll n$$

As $C_t - C'_t$ is small for $t \in I$, one expects that X'_{n+j} yields a good approximation for X_{n+j}. The approximative chain (X'_t) is Markovian and uniformly ergodic for $t \in I$, whence its distribution converges almost to π during I. This gives (i). Similarly, it can be shown that it 'forgets' most of the past during I, which gives (ii).

The proof of the ergodicity of the AM algorithm contains some restrictions. The assumptions of π being bounded with bounded support are in practice often fulfilled. In most cases we can approximate π using a target distribution with bounded support: the likelihood function is typically such that it decays rapidly or we can assume that our prior distribution has a bounded support. Nevertheless, the removal of the restrictions in Theorem 2 is an ongoing research (see also Sec. 3.6).

The adaptation technique of AM is demonstrated in Fig. 3.1 where the evolution of the proposal distributions are shown. The example is a 2-dimensional
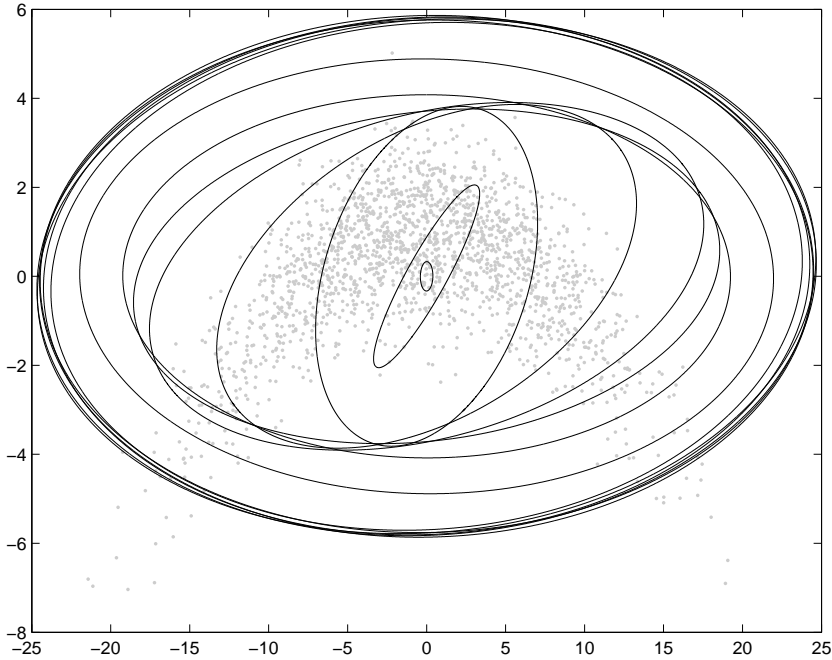
FIGURE 3.1. Demonstration of the AM algorithm: the evolution of the proposal distribution in a two-dimensional 'banana-shaped' test case. The proposal distributions (only 14 of them) are shown as ellipsoids representing the area covering 90% of the probability mass of the proposal distribution. The initial proposal distribution was the smallest ellipsoid and the last ones are the largest. The ellipsoids are centered at the origin to make them more easily comparable. The sampled points are indicated with light gray dots.

'banana-shaped' target distribution. In practice, the usefulness of the AM algorithm is based on the idea that the proposal distribution converges approximately to the (scaled) covariance matrix of the target distribution. Note also that it is not necessary to use the whole chain for the adaptation but suitably increasing part of it.

Most of the adaptive algorithms proposed so far are closely problem specific or very general in the sense that they discuss more about the setup in which adaptation could take place without proposing reasonably practical adaptation techniques. In this context the advantages of the AM algorithm are the following: (1) AM is simple, (2) it is easy to implement, (3) it is fast: the recursive formula for computing the covariance matrix can be used and the extra computational burden does not increase with time, (4) memory requirements are low (at least when the dimension is not too high) and do not increase with time and (5) it is automatic and therefore easy to use. The pseudo-code of the AM algorithm is given in Fig. 3.2.

AM ALGORITHM

1: Initialize: X^0 and $\overline{X}^0 = X^0$
2: Initialize: C^0 and $K^1 = C^0$
3: Initialize t^0
4: $f_{\text{new}} = p_{\text{pr}}(X^0)p(y \mid X^0)$.
5: $f_{\text{old}} = f_{\text{new}}$
6: FOR $t \leftarrow 1, \ldots, N$ DO
7:    IF $t < t^0$ THEN
8:       $C^t = C^0$
9:    ELSE
       $C^t = K^t + \varepsilon I_d$
10:    END IF
11:    $H^t = \text{Chol}(C^t)$ (Cholesky decomposition)
12:    Sample $G = [g_1, \ldots, g_n]^T$, where $g_i \sim \mathbf{N}(0,1)$ (Normal distribution).
13:    $Z = X^{t-1} + \sqrt{s_n}H^tG$.
14:    $f_{\text{new}} = p_{\text{pr}}(Z)p(y \mid Z)$.
15:    IF $f_{\text{new}} > f_{\text{old}}$ THEN
16:       $X^t = Z$
17:       $f_{\text{old}} = f_{\text{new}}$
18:    ELSE
19:       Sample s from uniform distribution $\mathbf{U}(0,1)$.
20:       IF $s < \frac{f_{\text{new}}}{f_{\text{old}}}$ THEN
21:          $X^t = Z$
22:          $f_{\text{old}} = f_{\text{new}}$
23:       ELSE
24:          $X^t = X^{t-1}$
25:       END IF
26:    END IF
27:    $\overline{X}^t = \frac{t}{t+1}\overline{X}^{t-1} + \frac{1}{t+1}X^t$.
28:    $K^{t+1} = \frac{t-1}{t}K^t + \overline{X}^{t-1}(\overline{X}^{t-1})^{\text{T}} - \frac{t+1}{t}\overline{X}^t(\overline{X}^t)^{\text{T}} + \frac{1}{t}X^t(X^t)^{\text{T}}$.
29: END FOR

FIGURE 3.2. Pseudo-code of the AM algorithm

3.5 THE SINGLE COMPONENT ADAPTIVE METROPOLIS ALGORITHM

When the dimension of the problem rises to a few hundreds, it is obvious that the sampling using AM becomes also slower. In high-dimensional problems the computation of the square root of the covariance matrix (i.e., Cholesky decomposition on line 11 in the pseudo-code) becomes simply more time consuming. Therefore, even if the covariance matrix is updated only at certain time intervals, the AM algorithm becomes slower. In high-dimensional problems the MCMC sampling is often realized by using the Gibbs sampling algorithm or the single component MH algorithm. The (to our knowledge) first fully adaptive MCMC algorithm that proceeds componentwise is Single Component Adaptive Metropolis algorithm (SCAM), introduced in PUBL. III. SCAM is a single component version

of the AM algorithm. The simple idea here is to update individually the variances of the one-dimensional Gaussian proposal distributions. The ergodicity proof of the SCAM algorithm follows the proof of AM [Haario, Saksman and Tamminen, 2003]. The natural requirement of the SCAM algorithm to work is that the target distribution is such that the standard single component algorithm (with Gaussian proposals) is ergodic.

Correlated target distributions are challenging for all single component methods and SCAM is not an exception in this sense. In such a situation some improvement can be achieved by rotating the sampling directions [e.g., Publ. III; Gilks and Roberts, 1996].

The SCAM algorithm is tested in varying dimensions up to 1000. The tests indicate that at least with rather well behaving target distributions the SCAM algorithm can be used in relatively high dimensions.

While MCMC techniques are nowadays used to solve really high-dimensional problems with tens of thousands of unknown parameters, it is important to keep in mind that the actual performance of MCMC in high-dimensional problems is still very much under research.

3.6 Variants and further development of the Adaptive Metropolis algorithm

In addition to the basic AM algorithm described above, some variants of the idea have been used in practice. In moderately high dimensions we have found it advisable to, instead of updating the covariance at every time, update it only at fixed time intervals. Another possibility is to weight the history differently to accelerate the freezing of the proposal distribution. It is obvious, that the effectiveness of the AM algorithm in its basic form is limited by the fact that the proposal distribution is chosen to be Gaussian. The idea behind the proof of Theorem 2 can be, however, applied to various other situations.

Andrieu and Robert [2001] extends the idea of non-Markovian adaptation with diminishing effect to a more general setup. Moreover, they make an interesting observation that the notions and techniques of stochastic approximation apply naturally in this context.

Andrieu and Moulines [2003] applies the techniques of stochastic approximation and proves the convergence of a modified AM algorithm to cover also targets with non-compact support. However, the density of the target is required to satisfy rather restrictive regularity conditions instead. An important outcome of their technique is the interesting estimate for the impact of the adaptation on the convergence speed. Indeed, in the case of standard AM algorithm their estimates show that the asymptotic error caused by the adaptation decays faster than the unavoidable error in MCMC corresponding to the central limit theorem.

Atchade and Rosenthal [2003] in turn generalizes the proof of Theorem 2 to the geometrically ergodic situation (whereas the original proof of Theorem 2 basically corresponds to the uniformly ergodic case). However, the conditions of Atchade and Rosenthal [2003] are fairly implicit and not very easy to verify in practice. They also propose an algorithm where the adaptation is based on monitoring the acceptance rate of the sampler and tune the size of the (spherical) Gaussian proposal distribution to achieve the optimal acceptance rate. Here they also observe the connection to stochastic approximation.

The idea of combining delayed rejection technique [Tierney and Mira, 1999] and the AM algorithm is introduced as DRAM algorithm in a recent preprint Haario, Laine, Mira and Saksman [2003]. It is shown that this combination may be especially helpful in some special situations where AM has problems in getting started.

4 Application: Atmospheric remote sensing by GOMOS satellite instrument

4.1 Motivation

Indirect remote sensing techniques are today routinely used for atmospheric research. The data processing of these instruments often involve solving nonlinear inverse problems. The traditional approach to solving such problems is to assume that the posterior distribution is Gaussian, at least around some maximum a posteriori (MAP) estimate, and to search for the MAP estimate either by linearizing the problem or using iterative optimization algorithms. The potential advantages of using MCMC for solving inverse problems are: (1) linearization is not needed, (2) freedom in implementing other than Gaussian prior information, (3) noise may be non-Gaussian, (4) modeling error can be taken into account in a flexible way, (5) getting trapped at local maxima is less probable than with optimization methods and (6) full characterization of the (non-Gaussian) posterior distribution is possible.

Earlier work related to different Monte Carlo methods that have been applied to geophysical problems include mainly inverse problems of Earth sciences, like seismology [e.g., recent reviews Mosegaard and Sambridge, 2002; Sambridge and Mosegaard, 2002]. In this work (Publ. IV–VI) we have studied the possibilities of using the MCMC technique in atmospheric remote sensing (see also Tamminen, Sihvola and Haario [1996]; Tamminen, Haario, Kyrölä and Oikarinen [1998]; Tamminen [1999]; Tamminen, Kyrölä and Auvinen [1999]; Auvinen, Oikarinen, Kyrölä, Tamminen and Leppelmeier [1999]). In particular, we have considered the inverse problems of the GOMOS satellite instrument.

4.2 GOMOS satellite instrument

GOMOS (Global Ozone Monitoring by Occultation on stars) is one of the 10 instruments onboard the European Space Agency's Envisat satellite (see Figure 4.1) which is targeted on studying the Earth's environment [ESA, 2001]. The Envisat satellite was launched from French Guyana on the 1st of March in 2002 to a polar, sun-synchronous orbit at about 800 km above the Earth. The main objective of GOMOS is to measure the atmospheric composition and especially the ozone concentration in the stratosphere and mesosphere with high vertical resolution [Bertaux, Hauchecorne, Dalaudier, Cot, Kyrölä, Fussen, Tamminen, Leppelmeier, Sofieva, Hassinen, d'Andon, Barrot, Mangin, Théodore, Guirlet, Korablev, Snoeij, Koopman and Fraisse, 2004; Kyrölä, Tamminen, Leppelmeier, Sofieva, Hassinen, Bertaux, Hauchecorne, Dalaudier, Cot, Korablev, d'Andon, Barrot, Mangin, Theodore, Guirlet, Etanchaud, Snoeij, Koopman, Saavedra, Fraisse, Fussen and
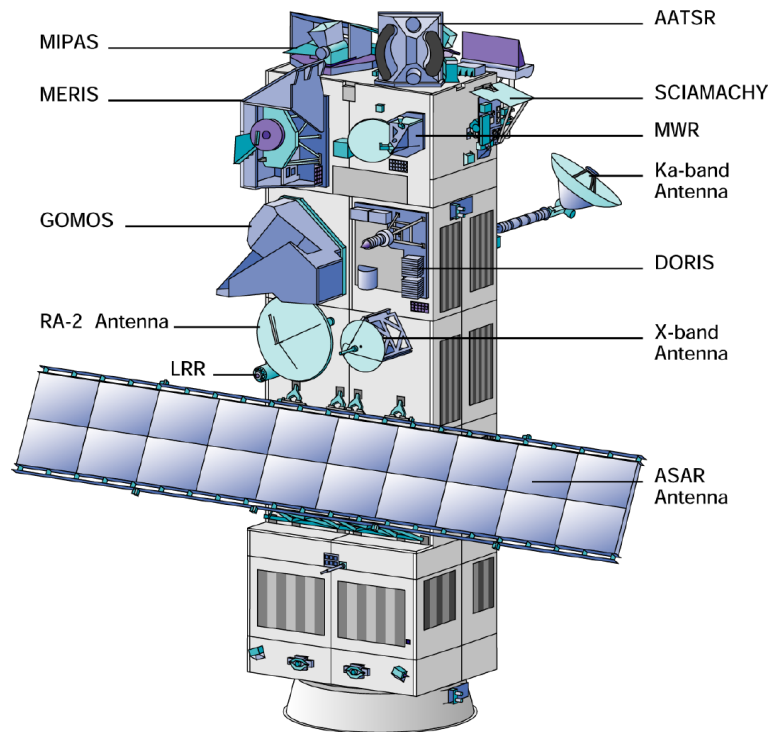
FIGURE 4.1. Envisat satellite and the instruments. In flight, GOMOS is looking backwards with respect to the direction of satellite velocity. (Figure provided by ESA).

Vanhellemont, 2004]. In addition to ozone (O_3) the UV-visible spectrometer (250–675 nm) can be used to detect also NO_2, NO_3, aerosols and neutral density. Two infra-red channels are used to detect O_2 and H_2O. The Finnish Meteorological Institute (FMI) has been involved in the GOMOS project right from the beginning; GOMOS was proposed together by FMI and the French Service d'Aeronomie in 1988 to ESA's Polar Platform satellite, which became later Envisat.

The GOMOS instrument is the first operational instrument that uses the stellar occultation technique to study the Earth's atmosphere [Bertaux, Megie, Widemann, Chassefiere, Pellinen, Kyrölä, Korpela and Simon, 1991]. The measurement principle, demonstrated in Figure 4.2, is elegant: the stellar spectrum seen through the atmosphere is compared with the reference spectrum measured above the atmosphere. Due to the absorption and scattering in the atmosphere the light measured through the atmosphere is attenuated and the attenuation is proportional to the amount of constituents in the atmosphere. The measurements are repeated at different tangential altitudes to obtain vertical profiles of the concentrations of different atmospheric constituents. The advantages of the GOMOS instrument compared to other instruments measuring ozone are the fairly good global coverage, with 300–400 occultations daily around the Earth (see Figure 4.3 for an example of the coverage of GOMOS occultations during one week in April
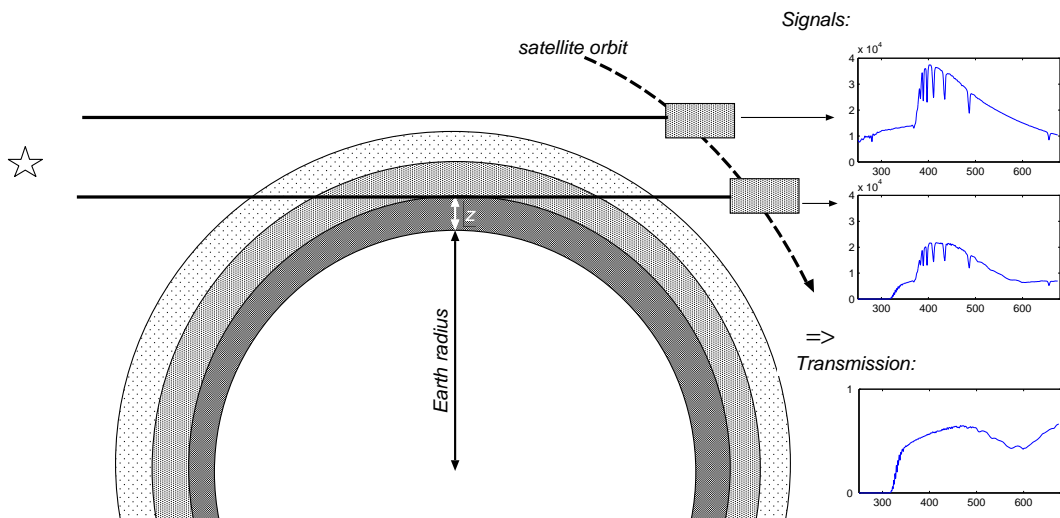
FIGURE 4.2. GOMOS measurement principle. The horizontal transmission of the atmosphere at tangent altitude z is obtained by dividing the attenuated stellar spectrum with the reference spectrum measured above the atmosphere.

2003) combined with the excellent vertical resolution (sampling resolution 0.3–1.7 km). The altitude range which can be covered by GOMOS is large: 15–100 km and the brightest stars can be followed even down to 5 km. Each occultation consists of about 70–100 spectra measured at different tangential altitudes and each UV-vis spectra includes measurements at 1416 different wavelengths. Due to the multitude of stars it is important that the optimal set of stars is selected for each orbit. This optimization is included in the GOMOS mission planning [Kyrölä and Tamminen, 1999].

4.3 GOMOS DATA RETRIEVAL

In the GOMOS data processing constituent densities are retrieved from stellar spectra attenuated in the atmosphere. The GOMOS inverse problem can be considered as an exterior problem in tomography [e.g., Natterer, 1986], but in practice it is solved locally considering only data collected from one occultation at a time. This inverse problem is as follows. By dividing the stellar spectrum measured through the atmosphere with the reference spectrum measured above the atmosphere we obtain a so called transmission spectrum. The transmission at wavelength λ, measured along the ray path ℓ, includes a term $T^{abs}_{\lambda,\ell}$ due to absorption and scattering by atmospheric constituents and a term $T^{ref}_{\lambda,\ell}$ due to refractive attenuation and scintillations, i.e., $T_{\lambda,\ell} = T^{abs}_{\lambda,\ell} T^{ref}_{\lambda,\ell}$. The dependence of the transmission on the constituent densities along the line of sight ℓ is given by
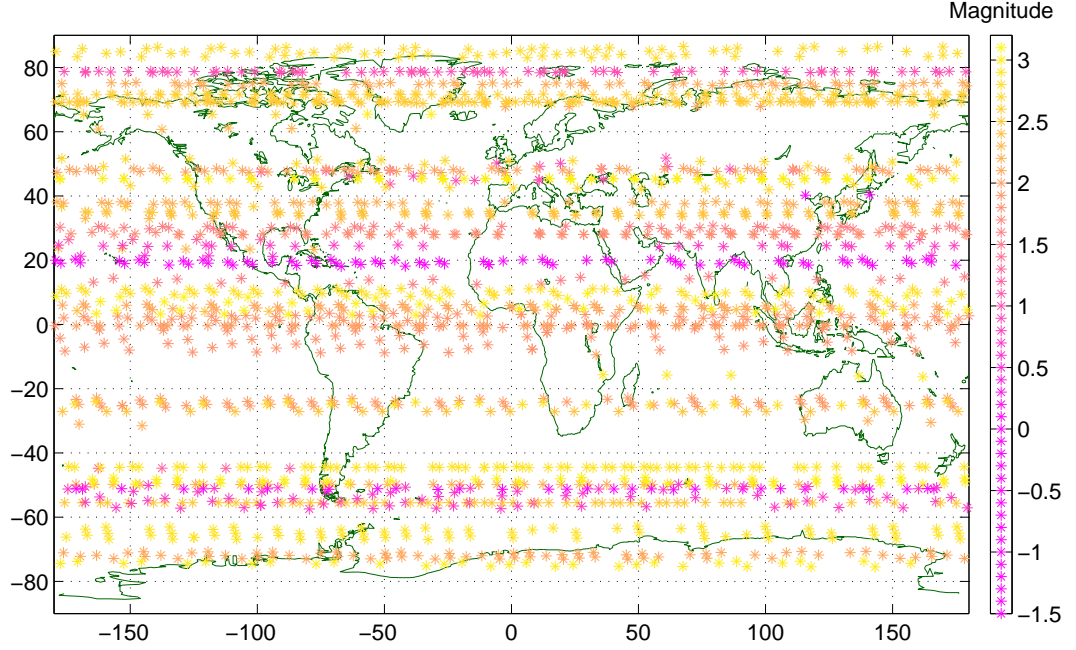
FIGURE 4.3. Geographical coverage of GOMOS measurements in one week (April 1–7, 2003). The gray scale (right panel) indicates the brightness (magnitude) of the star: bright stars are shown with pink and dim stars with yellow.

the Beer's law [e.g., Stephens, 1994]):

$$T_{\lambda,\ell}^{\mathrm{abs}} = \exp\left[-\int_{\ell} \sum_{\mathrm{gas}} \alpha_{\lambda}^{\mathrm{gas}}(z(s))\rho^{\mathrm{gas}}(z(s))ds\right],$$

where $\rho^{gas}(z)$ gives the constituent density at altitude z and α denotes the cross sections. Each atmospheric constituent has typical wavelength ranges where the constituent is active either by absorbing, scattering or emitting light. The cross sections reflect this behavior and their values are considered to be known from laboratory measurements. In the equation above the sum is over different gases and the integral is taken over the ray path. The problem is ill-posed in the sense that continuous profile is retrieved from a discrete set of measurements. Therefore some additional regularization or prior information is required to make the problem well-posed and solvable. In practice this is done by discretizing the atmosphere into layers and assuming, e.g., constant or linearly varying density inside layers. In PUBL. VI the problem of regularization is shortly mentioned, but the optimal amount of smoothness is an ongoing research [Tamminen, Kyrölä and Sofieva, 2004; Sofieva, Tamminen, Haario, Kyrölä and Lehtinen, 2004; Sofieva, Kyrö and Kyrölä, 2004].

The measurements are modeled by

$$y_{\lambda,\ell} = T_{\lambda,\ell}^{\mathrm{abs}} T_{\lambda,\ell}^{\mathrm{ref}} + \epsilon_{\lambda,\ell,}$$

assuming additive independent Gaussian noise $\epsilon_{\lambda,\ell} \sim \mathbf{N}(0, \sigma_{\lambda,\ell}^2)$, $\lambda = \lambda_1, \ldots, \lambda_\Lambda$, $\ell = \ell_1, \ldots, \ell_M$. The likelihood function for the constituent profiles then reads as

$$P(y \,|\, \rho(z)) \propto e^{-\frac{1}{2}(T-y)C^{-1}(T-y)}$$

with $C = \operatorname{diag}(\sigma_{\lambda,\ell}^2)$ and $y = (y_{\lambda,\ell})$, $T = (T_{\lambda,\ell})$. The inverse problem is to estimate the constituent profiles $\rho(z) = (\rho^{\text{gas}}(z))$, gas $= 1, \ldots, n_{\text{gas}}$.

In the operational data processing of GOMOS the problem is divided into two parts. The separation is possible if the measurement noise is independent between successive altitudes and the temperature-dependent cross sections can be sufficiently well approximated with 'representative' cross sections (e.g., cross sections at the temperature of the tangent point of the ray path) [Kyrölä, Sihvola, Kotivuori, Tikka, Tuomi and Haario, 1993; Sihvola, 1994]. In the operational algorithm these simplifications are assumed and the problem is solved in two steps. The spectral inversion is

$$T_{\lambda,\ell}^{\text{abs}} = \exp\left[-\sum_{\text{gas}} \alpha_{\lambda,\ell}^{\text{gas}} N_\ell^{\text{gas}}\right], \quad \lambda = \lambda_1, \ldots, \lambda_\Lambda,$$

which is solved for the horizontally integrated line-of-sight densities N_ℓ^{gas}. The vertical inversion

$$N_\ell^{\text{gas}} = \int_\ell \rho^{\text{gas}}(z(s))ds, \quad \ell = \ell_1, \ldots, \ell_M.$$

is solved for local constituent densities ρ^{gas} using the line-of-sight densities as the data. Note, that it is also possible to solve the problem directly in one step by inverting the local densities from the transmission data. This approach is here referred as the one-step inversion.

4.4   Implementing MCMC

Let us consider first the operational GOMOS data processing approach that consists of two steps: spectral inversion and vertical inversion. The spectral inversion problem is nonlinear and therefore a potential advantage may be obtained if it is solved using the MCMC technique. The dimension of the problem is small, only about 5 parameters (horizontally integrated line-of-sight densities of different constituents) to be retrieved but the inversion is done repeatedly at each altitude about 70-100 times for each occultation. The natural way of implementing the MCMC technique to such a problem is to use random walk MH algorithm since the posterior distributions are unknown and the Gibbs sampling or independence sampling MCMC algorithms can not be applied in a straightforward manner. Since the size of the problem is small and the posterior distributions are correlated multidimensional sampling is considered.
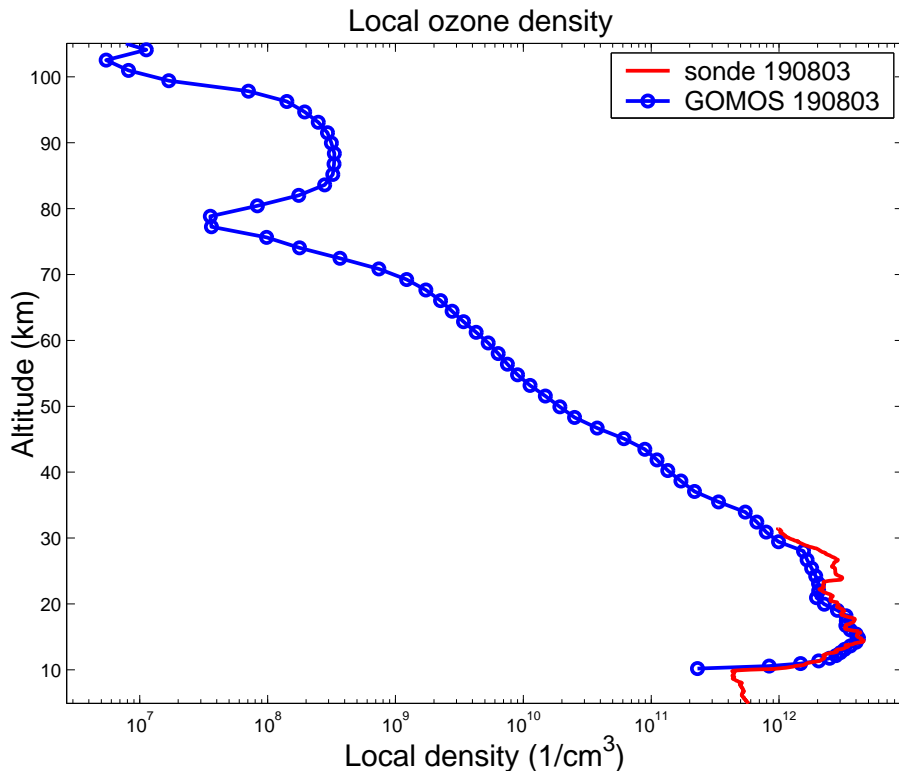
FIGURE 4.4. Ozone profiles measured by GOMOS and by an ozone sonde close to Marambio in Antarctica on August 19th 2003.

As discussed earlier, the main difficulties in implementing MH type MCMC to real inverse problems are typically related to tuning the proposal distribution to obtain efficient sampling. The special feature in the GOMOS data processing is that the posterior distributions of the spectral inversion vary strongly. They depend on the tangential altitude and also on the star used for the occultation. The line-of-sight densities vary typically several decades between 15 to 100 km (see Figure 4.4 for ozone vertical profile measured by GOMOS and by an ozone sonde). When the star is dim (and hence the signal-to-noise ratio is low) the posterior distributions become many times wider compared to the ones obtained for a bright star. In such a setup it is impossible to find any fixed proposal distribution that would work at all altitudes and for all stars. Therefore, the proposal distributions need to be optimized for each altitude and for each occultation separately. However, the manual tuning of the proposal distributions is also impossible to realize because of the huge number of different cases. Automatic algorithms for tuning the proposal distribution are therefore needed.

To overcome these problems we have applied the adaptive MCMC algorithms AP (PUBL. IV) and AM (PUBL. V–VI) to the GOMOS spectral inversion problems. To the GOMOS one-step problem we have applied the SCAM (PUBL. VI)

algorithm. The advantage of these algorithms is that they make the implementation of the MCMC easy: the adaptation can be used in a fully automatic way without increasing the computational time dramatically. The posterior distributions of GOMOS are not too strongly nonlinear (with multiple modes, for example) and therefore the inverse problems are suitable for the automatic algorithms.

4.5 IMPROVING AND VALIDATING GOMOS INVERSE ALGORITHMS WITH MCMC

The constituent profiles measured by GOMOS are further used in trend analysis, climatologies and in data assimilation. All of them rely strongly on the correctness of the error estimation of the GOMOS data. The proper statistical analysis and error characterization of GOMOS results is therefore important. In atmospheric remote sensing problems the posterior distributions have so far been approximated only with Gaussian distributions. The MCMC technique is a relatively efficient and easily applicable method for computing posterior distributions other than Gaussian. Therefore, it allows proper statistical analysis also for non-Gaussian problems like GOMOS inverse problem.

The advantages of the MCMC technique in the GOMOS data retrieval are demonstrated with simulated data in PUBL. IV and PUBL. VI. The data was simulated using FMI's relatively realistic GOMOS simulator called LIMBO [Kyrölä, Tamminen, Oikarinen, Sihvola, Verronen and Leppelmeier, 1999]. The ozone profile inverted from real GOMOS measurements using AM algorithm in the spectral inversion is shown in Fig. 4.5.

The operational data processing of satellite data is typically based on numerous assumptions to make the algorithms fast. With traditional techniques that are based on assuming normally distributed variables the validation of these assumptions may be difficult, whereas the MCMC technique provides relatively simple environment to test different assumptions. This is demonstrated in PUBL. V by using real GOMOS data from two occultations measured at 24th September in 2002. The AM algorithm turned out to work well in the problems considered. The constituent profiles retrieved with AM agree well with the profiles retrieved with the traditional methods used in the operational data processing applying the same assumptions as in the operational data processing. However, in some cases differences were observed. The flexibility of the MCMC technique to include other than Gaussian measurement noise or prior information can be used to improve the retrieval results. Constraints, like positivity, which are typically not used correctly as prior information in inverse problems as they lead to non-Gaussian posterior distributions, can easily bee implemented by the MCMC technique. In the GOMOS inversion the positivity prior and robust ℓ_1 norm solution seemed to improve the results in some cases. It is also shown that more realistic error
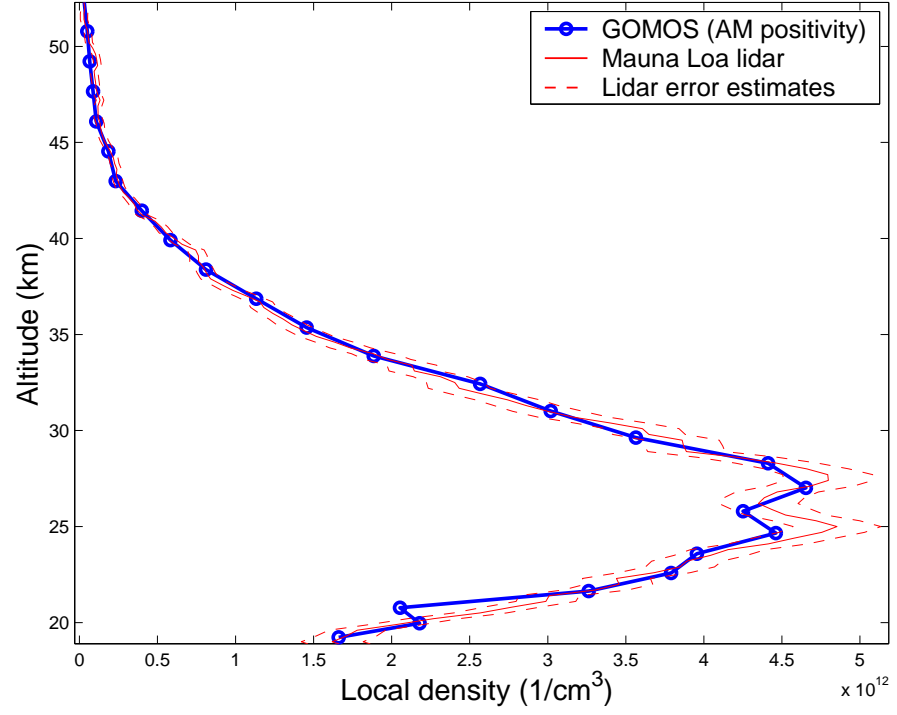
FIGURE 4.5. Ozone profiles measured by GOMOS (heavy solid line with circles) and an ozone lidar at Mauna Loa, Hawaii (thin solid line), on September 24th 2002. The error estimates ($1\sigma = 68.3\%$) of the lidar measurement are shown as dashed lines around the measurement.

estimates are obtained if modeling error is taken into account.

Finally, the full posterior distribution of GOMOS inverse problem can be obtained rather efficiently by solving the one-step inversion with the SCAM algorithm [PUBL. VI]. This approach does not rely on the fact that the measurement noise is independent at successive altitudes and the temperature dependence of the cross sections can be modeled in a proper way.

5 CONCLUDING REMARKS

The numerous adaptive MCMC algorithms proposed so far are tailored to over-come some specific, strongly target-related problems. In this work we have developed new adaptive MCMC algorithms that are by definition extremely simple to implement. New ideas and techniques have also been developed to treat the theoretical basis of the AM algorithm. The extensive testing of AM and SCAM shows that at least in the test examples used we obtain better estimates of target distribution than with corresponding non-adaptive algorithms with some 'non-optimal' proposal distributions. Indeed, the tests showed that for Gaussian targets the performance of AM and SCAM algorithms was equal with the 'optimally' tuned non-adaptive algorithms. It is also demonstrated that these algorithms can be used to solve real problems.

The main motivation for applying MCMC to GOMOS inverse problems is to improve the GOMOS inversion results by detecting possible pitfalls in the operational data processing and by validating the error estimates given by the operational processing. Because the satellite instruments are very expensive it is reasonable to use also state-of-the-art inversion methods to obtain as much information as possible from the data. In this context the MCMC technique is especially suitable for studying the error structure of the retrieved parameters, which is essential for further research, like ozone trend analysis or assimilating GOMOS measurements to atmospheric models.

PUBL. VI was presented at the Royal Statistical Society's meeting on 'Statistical approaches to inverse problems'. In the discussion that followed the presentation several remarks were made and we refer to Andrieu [2004]; Robert [2004]; Haario, Laine, Lehtinen, Saksman and Tamminen [2004] for further discussion.

Summaries of the original publications

I H. Haario, E. Saksman and J. Tamminen, (1999): Adaptive proposal distribution for random walk Metropolis algorithm. *Computational Statistics*, 14, 375–395.

Publ. I discusses the tuning of the size and the shape of the Gaussian proposal distribution of the random walk Metropolis algorithm. As a remedy to the laborious and time consuming manual tuning an automatic technique, called Adaptive Proposal (AP) algorithm, is introduced. The AP algorithm is based on updating the covariance matrix of the proposal distribution according to the empirical covariance of some fixed number of previously sampled points. The AP algorithm is simple and easy to implement and it is shown that in many cases the performance of the AP algorithm is sufficient. However, as the adaptation destroys the Markovian property of the chain, convergence to the correct target distribution is not guaranteed if the adaptation is continued after the burn-in phase. This is demonstrated with a 'tricky' 2-dimensional target distribution, where we find a clear discrepancy between the true target distribution and the one sampled with the AP algorithm by adapting continuously. In the publication we also introduce an extensive test procedure which has been used to test the performance of the AP algorithm and to compare it with the traditional Metropolis algorithm with different proposal distributions. As a realistic application the algorithm was successfully applied to a simulated GOMOS spectral inversion problem.

II H. Haario, E. Saksman and J. Tamminen, (2001): An adaptive Metropolis algorithm. *Bernoulli*, 7(2), 223–242.

In Publ. II we discuss further the tuning problem in Metropolis algorithm and introduce a non-Markovian adaptive MCMC algorithm called Adaptive Metropolis (AM) algorithm. The AM algorithm is similar to the AP algorithm: the covariance of the Gaussian proposal is updated by using the information of all (or suitably increasing part of) previously sampled points. In contrast to the AP algorithm the AM algorithm is shown to be ergodic in the sense that the law of large numbers holds. The adaptation can therefore be continued throughout the sampling. The ergodicity of the AM algorithm is also demonstrated with several test cases including the same 'tricky' 2-dimensional test which indicated problems in AP algorithm.

III H. Haario, E. Saksman and J. Tamminen, (2004): Componentwise adaptation for high dimensional MCMC. *Computational Statistics*, Accepted.

Publ. III discusses the MCMC sampling in relatively high-dimensional problems. In this publication a Single Component Adaptive Metropolis

(SCAM) algorithm is introduced. It combines the ideas of the AM algorithm and the single component sampling. The sampled chain is correctly ergodic and it is also shown to work well in practice up to 1000-dimensional problems (at least when the target distribution is relatively well behaving).

IV   J. Tamminen and E. Kyrölä, (2001): Bayesian solution for nonlinear and non-Gaussian inverse problems by Markov chain Monte Carlo method. *Journal of Geophysical Research*, 106(D13), 14,377–14,390.

In PUBL. IV the MCMC technique is discussed in the context of geophysical inverse problems. As an example, the GOMOS spectral inversion problem is solved by applying the AP algorithm. The potential advantage of using MCMC is demonstrated by different examples. For example, when the spectral inversion is linearized, the originally Gaussian noise becomes non-Gaussian. With the MCMC algorithm it is possible to take into account non-normal noise structure. The results are compared to the ones where the noise is approximated to be Gaussian. The improvement is clear when the noise is correctly modeled. The non-linear spectral inversion is nominally solved by iteratively searching the maximum point of the posterior distribution. With MCMC algorithm we can compute the expectation value of the posterior distribution. When the measurement data is noisy, the behavior of the iterative algorithm suffers and the solution is no more robust. In these cases the expectation of the posterior distribution seems to be the more robust estimator. In this publication the positivity prior is also discussed and it is shown to improve the results, especially when the signal-to-noise ratio is low.

V   J. Tamminen, (2004): Nonlinear inverse algorithm validation with Markov chain Monte Carlo. *Journal of Geophysical Research*, Accepted.

PUBL. V discusses how the MCMC technique can be used in the validation of operational data processing of atmospheric remote sensing measurements. We apply the MCMC methodology and especially the AM algorithm to the GOMOS spectral inversion and validate the operational inversion using real data measured by GOMOS. It is shown that when no prior information is used and the noise is assumed to be normally distributed the MCMC inversion and the fast operational algorithm give similar results. However, in some cases the results are improved when positivity or robust ℓ_1-norm is used. The inclusion of modeling error using the MCMC technique is also investigated.

VI   H. Haario, M. Laine, M. Lehtinen, E. Saksman and J. Tamminen, (2004): MCMC methods for high dimensional inversion in remote sensing, with discussion. *Journal of the Royal Statistical Society B*, 66, Part 3, 591–607.

In PUBL. VI the need and the advantage of the adaptive MCMC is discussed in the context of high-dimensional GOMOS inverse problems. Two approaches are introduced: the parallel method, applying the AM algorithm for sub-problems, which can be used to validate the GOMOS operational algorithms and the one-step inversion, applying the SCAM algorithm, which is free of some assumptions needed in the operational approach. In the publication discretization and regularization issues are also studied.

References

Andrieu, C. [2004], 'Discussion on the meeting on 'Statistical approaches to inverse problems'', *Journal of the Royal Statistical Society B* **66**(Part 3), 627–652.

Andrieu, C. and Moulines, E. [2003], On the ergodicity properties of some adaptive MCMC algorithms. Preprint.
URL: http://www.statslab.cam.ac.uk/mcmc/

Andrieu, C. and Robert, C. P. [2001], Controlled MCMC for optimal sampling. Preprint.
URL: http://www.statslab.cam.ac.uk/mcmc/

Atchade, Y. F. and Rosenthal, J. S. [2003], On Adaptive Markov Chain Monte Carlo Algorithms. Preprint.
URL: http://www.statslab.cam.ac.uk/mcmc/

Auvinen, H., Oikarinen, L., Kyrölä, E., Tamminen, J. and Leppelmeier, G. W. [1999], Inversion algorithms for OSIRIS and GOMOS bright-limb background term, *in* 'ESAMS99, European Symposium on Atmospheric Measurements from Space', Vol. WPP-161, pp. 271–276.

Bertaux, J. L., Hauchecorne, A., Dalaudier, F., Cot, C., Kyrölä, E., Fussen, D., Tamminen, J., Leppelmeier, G. W., Sofieva, V., Hassinen, S., d'Andon, O. F., Barrot, G., Mangin, A., Théodore, B., Guirlet, M., Korablev, O., Snoeij, P., Koopman, R. and Fraisse, R. [2004], 'First results on GOMOS/Envisat', *Advances in Space Research* **33**, 1029–1035.

Bertaux, J. L., Megie, G., Widemann, T., Chassefiere, E., Pellinen, R., Kyrölä, E., Korpela, S. and Simon, P. [1991], 'Monitoring of ozone trend by stellar occultations: The GOMOS instrument', *Advances in Space Research* **11**(3), 237–242.

Besag, J. and Green, P. J. [1993], 'Spatial statistics and Bayesian computation', *Journal of the Royal Statistical Society B* **55**(1), 25–37.

Chauveau, D. and Vandekerkhove, P. [2002], 'Improving convergence of the Hastings–Metropolis algorithm with an adaptive proposal', *Scandinavian Journal of Statistics* **29**(1), 13–29.

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

Erland, S. [2003], *On Adaptivity and Eigen-Decompositions of Markov Chains*, Norwegian University of Science and Technology, Trondheim, Norway. Dr.Ing. Thesis, 2003:102.

ESA [2001], *Envisat-GOMOS, An instrument for global atmospheric ozone monitoring*, Vol. SP-1244, European Space Agency.

Gåsemyr, J. [2003], 'On an adaptive version of the Metropolis–Hastings algorithm with independent proposal distribution', *Scandinavian Journal of Statistics* **30**(1), 159–173.

Gelfand, A. E. and Sahu, S. K. [1994], 'On Markov chain Monte Carlo acceleration', *Journal of Computational and Graphical Statistcs* **3**(3), 261–276.

Gelfand, A. E. and Smith, A. F. M. [1990], 'Sampling-based approaches to calculate marginal densities', *Journal of the American Statistical Association* **85**, 853–409.

Gelman, A. G., Roberts, G. O. and Gilks, W. R. [1996], Efficient Metropolis jumping rules, *in* J. M. Bernardo, J. O. Berger, A. F. David and A. F. M. Smith, eds, 'Bayesian Statistics V', Oxford Univ. Press, New York, pp. 599–608.

Geman, S. and Geman, D. [1984], 'Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images', *IEEE Transaction on Pattern Analysis and Machine Intelligence* **6**, 721–741.

Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. [1996], Introducing Markov chain Monte Carlo, *in* W. R. Gilks, S. Richardson and D. J. Spiegelhalter, eds, 'Markov Chain Monte Carlo in Practice', Chapman & Hall, New York, pp. 1–19.

Gilks, W. R. and Roberts, G. O. [1996], Strategies for improving MCMC, *in* W. R. Gilks, S. Richardson and D. J. Spiegelhalter, eds, 'Markov Chain Monte Carlo in Practice', Chapman & Hall, New York.

Gilks, W. R., Roberts, G. O. and George, E. I. [1994], 'Adaptive direction sampling', *Statistical Science* **43**, 179–189.

Gilks, W., Roberts, G. and Sahu, S. [1998], 'Adaptive Markov chain Monte Carlo through regeneration', *Journal of the American Statistical Association* **93**, 1045–1054.

Haario, H., Laine, M., Lehtinen, M., Saksman, E. and Tamminen, J. [2004], 'Discussion on the meeting on 'Statistical approaches to inverse problems'', *Journal of the Royal Statistical Society B* **66**(Part 3), 627–652.

Haario, H., Laine, M., Mira, A. and Saksman, E. [2003], DRAM: Efficient adaptive MCMC. Reports of the Department of Mathematics, University of Helsinki, Preprint 374.

Haario, H., Saksman, E. and Tamminen, J. [1998], Adaptive proposal distribution for random walk Metropolis algorithm. Reports of the Department of Mathematics, University of Helsinki, Preprint 176.

Haario, H., Saksman, E. and Tamminen, J. [2003], Componentwise adaptation for MCMC. Reports of the Department of Mathematics, University of Helsinki, Preprint 342.

Hammersley, J. M. and Handscomb, D. C., eds [1964], *Monte Carlo Methods*, Chapman and Hall, New York.

Hanson, K. M. and Cunningham, G. S. [1998], Posterior sampling with improved efficiency, *in* K. M. Hanson, ed., 'Proc. SPIE, Medical Imaging: Image Processing', Vol. 3338, pp. 371–382.

Hastings, W. [1970], 'Monte Carlo sampling methods using Markov chains and their applications', *Biometrika* **57**, 97–109.

Holden, L. [1998], Adaptive chains. Technical report SAND/11/98. Norwegian Computing Center.
URL: http://publications.nr.no/ad8.pdf

Kyrölä, E., Sihvola, E., Kotivuori, Y., Tikka, M., Tuomi, T. and Haario, H. [1993], 'Inverse theory for occultation measurements, 1, Spectral inversion', *Journal of Geophysical Research* **98**, 7367–7381.

Kyrölä, E. and Tamminen, J. [1999], GOMOS mission planning, *in* 'ESAMS99, European Symposium on Atmospheric Measurements from Space', Vol. WPP-161, ESA, Noordwijk, pp. 101–110.

Kyrölä, E., Tamminen, J., Leppelmeier, G. W., Sofieva, V., Hassinen, S., Bertaux, J.-L., Hauchecorne, A., Dalaudier, F., Cot, C., Korablev, O., d'Andon, O. F., Barrot, G., Mangin, A., Theodore, B., Guirlet, M., Etanchaud, F., Snoeij, P., Koopman, R., Saavedra, L., Fraisse, R., Fussen, D. and Vanhellemont, F. [2004], 'GOMOS on Envisat: An overview', *Advances in Space Research* **33**, 1020–1028.

Kyrölä, E., Tamminen, J., Oikarinen, L., Sihvola, E., Verronen, P. and Leppelmeier, G. W. [1999], LIMBO—Limb and occultation measurement simulator, *in* 'ESAMS99, European Symposium on Atmospheric Measurements from Space', Vol. WPP-161, ESA, Noordwijk, pp. 487–493.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. [1953], 'Equations of state calculations by fast computing machine', *Journal of Chemical Physics* **21**, 1087–1091.

Mosegaard, K. and Sambridge, M. [2002], 'Monte Carlo analysis of inverse problems', *Inverse Problems* **18**, R29–R54.

Mykland, P., Tierney, L. and Yu, B. [1995], 'Regeneration in Markov chain samplers', *Journal of the American Statistical Association* **90**, 233–241.

Natterer, F. [1986], *The mathematics of computerized tomography*, John Wiley & Sons Ltd, Great Britain.

Nummelin, E. [1984], *General irreducible Markov chains and non-negative operators*, Cambridge University Press, Cambridge.

Nummelin, E. [2002], 'MC's for MCMC'ists', *International Statistical Review* **70**(2), 215–240.

Robert, C. P. [2004], 'Discussion on the meeting on 'Statistical approaches to inverse problems'', *Journal of the Royal Statistical Society B* **66**(Part 3), 627–652.

Robert, C. P. and Casella, G. [2000], *Monte Carlo statistical methods*, Springer.

Sahu, S. K. and Zhigljavsky, A. A. [2003], 'Self regenerative Markov chain Monte Carlo with adaptation', *Bernoulli* **9**, 395–422.

Sambridge, M. and Mosegaard, K. [2002], 'Monte Carlo methods in geophysical inverse problems', *Reviews of Geophysics* **40(3)**.

Sihvola, E. [1994], *Coupling of spectral and vertical inversion in the analysis of stellar occultation data*, Geophysical publications, no. 38, Finnish Meteorological Institute, Helsinki. Licentiate thesis at the University of Helsinki, Department of Theoretical Physics.

Sofieva, V. F., Kyrö, E. and Kyrölä, E. [2004], Smoothness of ozone profiles: Analysis of 11-years ozone sonde measurements at Sodankylä. Annales Geophysicae, accepted.

Sofieva, V. F., Tamminen, J., Haario, H., Kyrölä, E. and Lehtinen, M. [2004], A priori smoothness on ozone profile smoothness in the inversion from limb measurements. Annales Geophysicae, accepted.

Sokal, A. D. [1989], Monte Carlo methods in statistical mechanics: Foundations and new algorithms. Lecture Notes, Cours de Troisieme Cycle de la Physicque en Suisse Romade.

Stephens, G. L. [1994], *Remote sensing of the lower atmosphere, an introduction*, Oxford University Press.

Tamminen, J. [1999], MCMC methods for inverse problems, Geophysical Publication 48, Finnish Meteorological Institute, Helsinki.

Tamminen, J., Haario, H., Kyrölä, E. and Oikarinen, L. [1998], Data processing of the GOMOS instrument by using an adaptive MCMC method, *in* W. L. Barnes, ed., 'SPIE Proceedings Vol. 3439, Earth Observing Systems III', SPIE, pp. 470–479.

Tamminen, J., Kyrölä, E. and Auvinen, H. [1999], MCMC algorithms for inverse problems in remote sensing, *in* 'ESAMS99, European Symposium on Atmospheric Measurements from Space', Vol. WPP-161, ESA.

Tamminen, J., Kyrölä, E. and Sofieva, V. [2004], Does prior information improve measurements?, *in* G. Kirchengast, U. Foelsche and A. Steiner, eds, 'Occultations for Probing Atmosphere and Climate - Science from the OPAC-1 Workshop', Springer Verlag, pp. 87–98.

Tamminen, J., Sihvola, E. and Haario, H. [1996], Data processing and sensitivity studies of the GOMOS instrument, *in* P. B. Hayes and J. Wang, eds, 'SPIE Proceedings Vol. 2830, optical Spectroscopic Techniques and Insrumentation for Atmospheric and Space Reseach II', SPIE, pp. 180–188.

Tierney, L. [1994], 'Markov chains for exploring posterior distributions, with discussion', *Annals of Statistics* **22**, 1701–1762.

Tierney, L. and Mira, A. [1999], 'Some adaptive Monte Carlo methods for Bayesian inference', *Statistics in Medicine* **18**, 2507–2515.

Warnes, G. R. [2001], The Normal Kernel Coupler: An adaptive Markov chain Monte Carlo method for efficiently sampling from multi-modal distributions. Technical Report no 39, Department of Statistics, University of Washington.