

Antonietta Mira

Ordering and improving Monte
Carlo Markov chains performance

2002/2



UNIVERSITÀ DELL'INSUBRIA
FACOLTÀ DI ECONOMIA

<http://eco.uninsubria.it>

In questi quaderni vengono pubblicati i lavori dei docenti della Facoltà di Economia dell'Università dell'Insubria. La pubblicazione di contributi di altri studiosi, che abbiano un rapporto didattico o scientifico stabile con la Facoltà, può essere proposta da un professore della Facoltà, dopo che il contributo sia stato discusso pubblicamente. Il nome del proponente è riportato in nota all'articolo. I punti di vista espressi nei quaderni della Facoltà di Economia riflettono unicamente le opinioni degli autori, e non rispecchiano necessariamente quelli della Facoltà di Economia dell'Università dell'Insubria.

These Working papers collect the work of the Faculty of Economics of the University of Insubria. The publication of work by other Authors can be proposed by a member of the Faculty, provided that the paper has been presented in public. The name of the proposer is reported in a footnote. The views expressed in the Working papers reflect the opinions of the Authors only, and not necessarily the ones of the Economics Faculty of the University of Insubria.

© Copyright Antonietta Mira
Printed in Italy in January 2002
Università degli Studi dell'Insubria
Via Ravasi 2, 21100 Varese, Italy

All rights reserved. No part of this paper may be reproduced in any form without permission of the Author.

Ordering and improving Monte Carlo Markov chains performance

Antonietta Mira *

January 15, 2002

Abstract

An overview of orderings defined on the space of Markov chains having a pre-specified distribution as their unique stationary distribution is provided. The intuition gained by studying these orderings is used to improve existing Markov chain Monte Carlo algorithms.

Key words and phrases: Asymptotic variance, Convergence ordering, Covariance ordering, Efficiency ordering, Markov chain Monte Carlo, Metropolis-Hastings algorithm, Peskun ordering.

1 Motivation

Suppose we are given a probability distribution π on a state space $\mathbf{X}$, with associated Borel σ -algebra $\mathcal{B}(\mathbf{X})$, and we are interested in estimating the expectation of some function f , $E_\pi f(X) = \mu$. The Markov chain Monte Carlo (MCMC) estimate of μ is the ergodic average $\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n f(X_i)$ where X_i s are dependent samples collected along the path of a Markov chain that has π as its unique stationary (and limiting) distribution. The Markov chain is identified with its transition kernel (a matrix in finite state spaces): $P(x, A) = \Pr(X_n \in A | X_{n-1} = x)$ where $A \in \mathcal{B}(\mathbf{X})$.

One of the key observations that motivates this paper is the fact that, given a distribution π , there are many Markov chains that are stationary with respect to it and could therefore be used for MCMC purposes. A choice is thus necessary and criteria to drive this choice are needed. Orderings defined on the space of Markov chain that have a specified stationary distribution allow a comparison and can aid the selection of one chain over another. The second motivation is the fact that often the intuition behind an ordering leads to the definition of new MCMC algorithms or to the improvement of existing ones. This twofold motivation drives the structure of the paper.

In this manuscript I will mainly refer to finite state spaces and reversible transition matrices and point out when extensions of the results stated are available (up to my knowledge) in different settings.

*Università dell'Insubria, Facoltà di Economia, Via Ravasi 2, 21100, Varese, Italy. Email: amira@eco.uninsubria.it

2 Ordering MCMC relative to their performance

The two main criteria to evaluate the performance of a transition matrix used for MCMC simulations are *asymptotic variance* (AV) of the resulting estimates and *speed of convergence* (SC) to stationarity (typically measured by total variation distance). In this paper I will mainly focus on AV.

As observed by Besag and Green (1993), the two goals lead to different notions of optimality, since AV depends on the eigenvalues and SC on the absolute values of the eigenvalues of the transition matrix. Thus, while the optimal sampler will be different, we could get good performance with respect to both criteria simultaneously. Still, a suggested practice is to use a sampler with good convergence properties for the first part of the simulation (until convergence is presumably achieved) and then switch to a sampler with better properties in terms of AV.

Only if the transition matrix is positive definite that is its eigenvalues are positive, the two goals of AV and SC are not in conflict. There are a few Markov chains with positive transition matrices used for MCMC purposes: the independence Metropolis-Hastings (MH) algorithm (Liu, 1996a), the random scan Gibbs sampler (Liu et al. 1995), and the slice sampler (Mira and Tierney, 2001).

2.1 Orderings relative to the speed of convergence to stationarity

To justify this criterion as a measure of performance for MCMC purposes consider the fact that π , the distribution of interest, is both the unique stationary and the limiting distribution of the Markov chain we choose to simulate. This means that if we could start our chain by generating an observation from π then we are guaranteed that the updating mechanism (i.e. the transition matrix P) preserves this distribution: $\pi P = \pi$. But if we could generate i.i.d. observations from π we would do standard Monte Carlo integration instead of MCMC. We thus start our chain from an arbitrary initial distribution and then rely on the fact that π is also the limiting distribution. This means that, as the simulation goes on ($n \rightarrow \infty$), the distribution of the chain at time n , $P^n(x, \cdot)$, becomes closer and closer to π in total variation distance: $\|P^n(x, \cdot) - \pi\| \rightarrow 0$ regardless of the starting point x , where $\|P^n(x, \cdot) - \pi(\cdot)\| = \sup_{A \in \mathcal{B}(\mathbf{X})} |P^n(x, A) - \pi(A)|$.

The rate of convergence is governed by the spectral radius, which, in finite state spaces is the second largest eigenvalue (of the transition matrix) in absolute value (Besag and Green, 1993 and Roberts, 1996, p. 48). This leads to the definition of a natural ordering. Let $\mathcal{S}$ be the class of Markov chains stationary with respect to π and $\mathcal{R}$ the subset of the reversible ones. Let $\sigma(P)$ be the spectrum of P : the set of λ 's such that $\lambda I - P$ is not invertible, where I denotes the identity matrix. The spectrum of a Markov chain transition matrix is a non-empty closed subset of the interval $[-1, +1]$ and always contains an eigenvalue equal to one associated to constant functions. For this reason we also define $\sigma_1(P)$ which equals $\sigma(P)$ except that the eigenvalue associated with constant functions is removed. Notice that the eigenvalues are real-valued only if the Markov chain is reversible. This is why in the

following definition the class $\mathcal{R}$ is considered.

Definition 1 *Let $P, Q \in \mathcal{R}$. P dominates Q in the convergence ordering, $P \succeq_{SC} Q$, if*

$$\sup\{|\lambda| : \lambda \in \sigma_1(P)\} \leq \sup\{|\lambda| : \lambda \in \sigma_1(Q)\}.$$

This ordering appears also in Frigessi et al. (1993) where the random scan, single site update, MH algorithm and Gibbs sampler are compared in terms of SC both in general and for simulating the Ising model at different temperatures. Besides Frigessi et al. (1993) there are only a few other papers available in the literature that refer to the convergence ordering due to the fact that studying the spectral structure of Markov chains used for practical applications is typically not an easy task. As a consequence the papers we reference in the sequel only consider special Markov chains such as the Gibbs sampler, the independence MH and the Slice sampler or study particular distribution of interest that are easier to analyse such as Gaussian distributions or the Ising model. We recall the following results. Given any independence Metropolis-Hastings sampler it is always possible to construct a slice sampler that dominates it in the convergence ordering (Mira and Tierney, 2001). Theoretical results on convergence properties of the independence Metropolis-Hastings sampler compared to various Monte Carlo algorithm can be found in Liu (1996a). Finally, in Roberts and Sahu (1997), many different types of Gibbs sampler on Gaussian distributions are compared in terms of SC. The work covers parameterisation, blocking, random and deterministic scan. The comparisons are partially extended to the non-Gaussian case via a weak convergence result in Roberts and Sahu (2001). The relationship in terms of SC between the EM algorithm and the Gibbs sampler is investigated in Sahu and Roberts (1999). The authors show that the rate of convergence of the Gibbs sampler, obtained by Gaussian approximation, is equal to that of the corresponding EM type algorithm.

The lack of theoretical results is more than compensated by numerous papers where different MCMC samplers are compared in terms of speed of convergence to stationarity via simulation studies.

Different distances could be used to measure how far $P^n(x, \cdot)$ is from $\pi(\cdot)$ such as the χ^2 distance (Diaconis et al. 2000). Furthermore other form of convergence could be used to evaluate the performance of a Markov chain for MCMC purposes. We can, for example, consider convergence of moments or pathwise convergence (Roberts and Tweedie, 2001). Depending on the focus of our interest different orderings concerning convergence of MCMC methods can be defined. We do not pursue this further.

2.2 Orderings relative to the asymptotic variance of the estimates

To justify the choice of this criterion consider that, in classical statistics, estimates are compared in terms of their asymptotic relative efficiency. If a Markov chain is irreducible and the state space is finite the corresponding MCMC estimates are strongly consistent and asymptotically normally distributed, therefore efficiency will

be measured by the asymptotic variance of estimates. Under slightly stronger conditions asymptotic normality and consistency are guaranteed on general state space as well (Roberts and Tweedie, 2001).

Let $v(f, P)$ be the limit, as n tends to infinity, of $n\text{Var}_\pi[\hat{\mu}_n]$ for a chain updated using the transition kernel P . Sometimes we run a MCMC simulation having in mind a specific function of interest whose expectation relative to π we want to estimate. In other situations we either have a range of functions in mind or we do not have any specific function and are interested in studying π per-se. In the first setting we will be interested in the choosing a sampler that has good efficiency properties *relative* to the specific function of interest. In the second setting we will look for a chain that is more efficient *uniformly* over all sensible f s. This distinction leads to the following definitions:

Definition 2 *Let $P, Q \in \mathcal{S}$. P is at least as efficient as Q relative to a particular function f , $P \succeq_f Q$, if $v(f, P) \leq v(f, Q)$.*

Definition 3 *Let $L^2(\pi)$ be the space of all f that have a finite variance with respect to π . Let $P, Q \in \mathcal{S}$. P is at least as uniformly efficient as Q , $P \succeq_E Q$, if $v(f, P) \leq v(f, Q)$ for all $f \in L^2(\pi)$.*

2.2.1 Uniform efficiency

The first ordering of MCMC samplers introduced in the literature refers to uniform efficiency and is due to Peskun (1973), a student of Hastings.

Definition 4 *Let $P, Q \in \mathcal{R}$. P dominates Q in the Peskun sense if each of the off-diagonal elements of P is greater than or equal to the corresponding off-diagonal elements of Q .*

Tierney (1998) extends Peskun ordering from finite to general state spaces. These orderings provide a sufficient condition for uniform efficiency as the following theorem (Peskun, 1973 and Tierney, 1998) states:

Theorem 5 *If P dominates Q in the Peskun sense then $P \succeq_E Q$ and $\lambda_{iP} \leq \lambda_{iQ}$ for all i where $\{\lambda_{0P} \geq \lambda_{1P} \geq \dots\}$ are the eigenvalues of P , arranged in decreasing order.*

The above theorem readily extends to general state spaces (Mira and Geyer, 1999) except that the eigenvalues cannot be ordered anymore (since it does not even make sense to talk about eigenvalues). What can be ordered are the respective suprema of the spectra of the operators defined by the transition kernels.

The first use of Peskun ordering appears in Peskun (1973) and states that the Metropolis-Hastings algorithm dominates a class of competitors reversible with respect to some π and all with the same propose/accept updating structure (with symmetric acceptance probability). An other interesting use of the ordering appears in Tierney (1998) where the author compares the benefit in terms of uniform efficiency of two approaches to using mixtures of MH kernels. In Frigessi et al.

(1992) the optimal transition matrix relative to the Peskun ordering is constructed. Unfortunately the construction only applies to finite state spaces and requires the exact knowledge of π while, in typical MCMC applications, π is known only up to a normalizing constant (this is one of the strengths of MCMC methods).

Unfortunately there are many Markov chains that are not comparable in the Peskun sense. To give just one example consider a transition matrix with zeros along the main diagonal. Since the row sums have to be equal to one Peskun will never be able to order them. This motivates the introduction, in Mira and Geyer (1999), of a weaker ordering, the covariance ordering:

Definition 6 Let $P, Q \in \mathcal{R}$. P dominates Q in the covariance sense, $P \succeq_C Q$, if $Q - P$ is a positive definite matrix.

Mira and Geyer (2000) extend this ordering from finite to general state spaces. As shown in Mira and Geyer (1999), the covariance ordering is equivalent to requiring that all the lag-one autocovariances are ordered that is

Theorem 7 Let $\text{Cov}_\pi(f, Qf) = E_\pi[f(X_0)f(X_1)]$ be the lag one autocovariance and $L_0^2(\pi)$ be the functions of $L^2(\pi)$ with zero mean relative to π . The condition $P \succeq_C Q$ is equivalent to $\text{Cov}_\pi(f, Qf) \geq \text{Cov}_\pi(f, Pf)$, for all $f \in L_0^2(\pi)$.

The covariance ordering does not imply the Peskun ordering, but the reverse implication is true. Hence the covariance ordering is a more general (weaker) criterion. Figure 1 illustrates how the concept of stationarity, reversibility, covariance order and Peskun order are related.

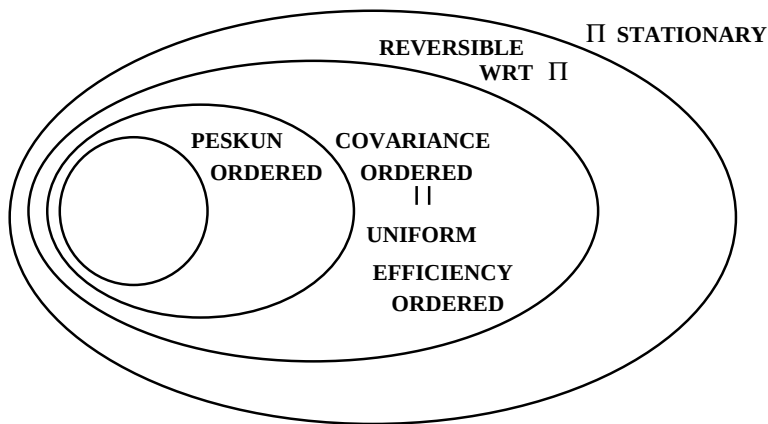


Figure 1: Relationships between MCMC related concepts.

The next theorem (Mira and Geyer, 1999) gives the main reason of interest in the covariance ordering:

Theorem 8 P is uniformly more efficient than Q if and only if P dominates Q in the covariance sense.

A necessary and sufficient condition for uniform efficiency in a non-reversible setting is given in Mira and Geyer (2000). At first sight Theorem 8 may seem a little

counter intuitive: by imposing a condition on the relative ordering of the lag one autocovariances we can order the asymptotic variance of $\hat{\mu}$ which, by definition, equals the double infinite sum of all lags autocovariances. The fact is that the condition imposed on the lag one autocovariances is stronger than it seems since it is required to hold over a big set of functions (namely $L_0^2(\pi)$).

2.2.2 Relative efficiency

Suppose now we have a specific function in mind f and, without loss of generality, assume it is monotone.

Let $\mathcal{X}$ be a finite set with N elements. Define the summation matrix T to be an $N \times N$ upper triangular matrix with zeros below the main diagonal and ones elsewhere. Define the south-west sub-matrix of the matrix M to be the sub-matrix of M obtained deleting the first row and the last column. Let Π be the $N \times N$ diagonal matrix with π_i as the i -th element on the diagonal. Indicate by f' the transpose of f .

An ordering which refers to relative efficiency appears in Mira (2000):

Definition 9 *Let $P, Q \in \mathcal{S}$. P dominates Q in the south-west ordering (with respect to some re-ordering of the state space that makes the function of interest monotone), $P \succeq_{sw} Q$, if all the elements of the south-west sub-matrix of $T\Pi(P - Q)T$ are non-negative.*

This ordering is interesting because of the following theorem (Mira, 2000):

Theorem 10 *Let P and Q be two Markov chains having π as their unique stationary distribution. Let $l_P^{-1} = (I - P + A)^{-1}$ be the inverse Laplacian. If $l_P^{-1} \succeq_{sw} l_Q^{-1}$ then $P \succeq_f Q$.*

A positive side of the south-west ordering is that it allows to compare also transition matrices that are not reversible unlike most of the other orderings introduced so far. On the other hand a limitation of Theorem 10 is that it requires to work with inverse Laplacians since computing the inverse can be computationally intensive. A way to work with the transition matrices directly is to realize that $(I + P - A)$ provides a first degree approximation to the inverse Laplacian (Peskun, 1973). This realization leads to the following

Definition 11 *Let $P, Q \in \mathcal{S}$. P is at least as first degree efficient as Q for a particular function f , $P \succeq_{1f} Q$, if $f'\Pi(P - Q)f \leq 0$.*

A reasoning similar to the one used in the proof of Theorem 10 shows that, if $P \succeq_{sw} Q$, then $P \succeq_{1f} Q$.

3 Improving the efficiency of MCMC samplers

3.1 Improving with respect to absolute efficiency

What can we learn from Peskun ordering? The intuition behind Theorem 5 is clear: a Markov chains that has smaller probability of remaining in the same position over

time (i.e. smaller diagonal elements in the transition matrix) explores the state space more efficiently and thus produces better estimates. This suggests that, in principle, we could improve the Metropolis-Hastings and the reversible jumps (Green, 1995) algorithms relative to the Peskun ordering by decreasing the rejection frequency of proposed moves. This is the intuition used in Liu (1996b) to improve the Gibbs sampler on discrete state spaces. A more general use of the same intuition appears in Tierney and Mira (1999) and Green and Mira (1999) where a delaying rejection mechanism is introduced to improve the MH and the Reversible Jumps algorithms (respectively) on a sweep by sweep basis.

The delaying rejection strategy works as follows: suppose that, at some point in time, the chain is at position x . Generate a candidate move y from some proposal distribution that may depend on the current position of the chain, $q_1(x, \cdot)$. With the usual MH probability $\alpha_1(x, y)$ accept the move. If the move is rejected, instead of staying put and advancing time which is what Metropolis, Hastings and Green would tell us to do (but we know this leads to a loss in terms of efficiency of the resulting MCMC estimates: Peskun told us!) propose a second stage candidate move z by generating it from a new proposal distribution $q_2(x, y, \cdot)$ and with some probability $\alpha_2(x, z)$ accept the move. If z is also rejected we could either interrupt the delaying rejection process and remain in the current position (advancing time) or continue with higher stages proposals. In Tierney and Mira (1999) and Green and Mira (1999) formulae are provided for second stage acceptance probabilities that preserve reversibility with respect to π separately at each stage of the delaying process. In Mira (2001) iterative formulae for higher stages acceptance probabilities are given. Notice that the second (and higher) stage proposal can be different from the first stage one and is allowed now to depend on the previously rejected candidate. This means that a form of within steps adaptation of the proposal distribution is allowed: we can learn from our previous “mistakes” (without loosing the Markovian property of the resulting sampler).

The delaying rejection strategy certainly reduces the overall probability of remaining in the current state and thus leads to improved samplers in the Peskun ordering. The price paid for this gain is the fact that it takes longer (in terms of simulation time) for the chain to perform a step (sweep). Thus, whether delaying rejection is useful in practice, depends on whether the reduction in variance obtained more than compensates for the additional computational cost. At this regard compare the results reported in Green and Mira (1999) where simulation time is taken into account. In all the examples reported the gain more than compensates the price.

3.2 Improving with respect to relative efficiency

Mira, (2000) introduces the idea of stationarity preserving and efficiency increasing probability mass transfers (S.P.T.) performed on a transition matrix P that has the proper stationary distribution. These S.P.T. are motivated by Definition 11 and are performed in the following way: given integers $1 \leq i < j \leq N$ and $1 \leq k < l \leq N$ and a quantity $h > 0$, increase $p_{i,l}$ and $p_{j,k}$ by h and $h\pi_i/\pi_j$ respectively and decrease $p_{i,k}$ and $p_{j,l}$ by h and $h\pi_i/\pi_j$ respectively. We note that the knowledge of π up

to a normalizing constant is sufficient to perform S.P.T.

If P is derived from Q via a sequence of S.P.T. then $P \succeq_{SW} Q$. We can thus increase the efficiency of a transition matrix (provided the first degree approximation to the asymptotic variance is good), while preserving its stationary distribution, via S.P.T. Indeed, we can keep moving probability mass around until there exist two indexes $i < j$ such that $p_{i,k}$ and $p_{j,l} > 0$ for some $k < l$.

The limiting transition matrix obtained by applying a sequence of S.P.T., that is a matrix that cannot be further improved by S.P.T. has at most one non zero element along the main diagonal and it presents a path of positive entries connecting the north-east to the south-west corner of the matrix. Which specific pattern is optimal depends on π and f . The intuition behind the structure of the limiting matrix is the following: when we perform S.P.T. we move mass from the main diagonal (or along the main diagonal) of the transition matrix to the off-diagonal entries. The way the probability mass is distributed is such that big jumps between values of $f_i \pi_i$ are favored. In other words we try to induce negative correlation along the chain path so that the variance of the resulting MCMC estimates will be reduced, possibly even to values smaller than the variance we would get with i.i.d. sampling from π .

If f does not assume the same value on any two points of the state space the final result of a sequence of S.P.T. is *unique* and corresponds to the matrix that minimises the linear function $f \Pi(P - A) f'$ under the set of linear constraints that ensure that the resulting matrix is stochastic and has the proper stationary distribution (these constraints defines a convex region thus the minimum is unique). We will refer to such matrix as to first degree f -optimal matrix, 1- f -opt-matrix in short. Furthermore the 1- f -opt-matrix is *reversible* with respect to the stationary distribution π even if the starting point is only stationary with respect to π .

As pointed out by Billera and Diaconis (2001), there is an insightful geometrical interpretation of how the Metropolis-Hastings algorithm transforms a generic stochastic matrix K , the proposal distribution, into a reversible Markov chain (indicated as MH in Figure 2). Along the same lines we can interpret the final result of a sequence of S.P.T. transfers. In the sequel we will use the same notation and parametrization of a stochastic matrix used in Billera and Diaconis (2001) and reported in Figure 2. Consider the case of a stationary distribution with $\pi_1 < \pi_2$ (a similar picture is obtained if $\pi_1 \geq \pi_2$). In Figure 2 we represent the MH transition matrix obtained starting from a generic proposal K together with the first degree optimal matrix (1- f -opt) that we would obtain by taking any reversible matrix (i.e. any matrix on the reversible line of equation $b = a\pi_1/\pi_2$) and applying a sequence of S.P.T.

The second largest eigenvalue λ_2 of a two by two matrix parametrized as in the figure is $\lambda_2 = 1 - a - b$. Figure 2 thus clearly shows that the first degree optimal transition matrix has a smaller second largest eigenvalue than the MH matrix.

Finally notice that the first degree optimal matrix has a second largest eigenvalue smaller than any other matrix reversible with respect to π . This is an indication of the optimal performance of the corresponding Markov chain in terms of the convergence ordering. But note that this is an artifact of the fact that we are considering a state space with only two degrees of freedom. Furthermore the 1- f -opt. matrix

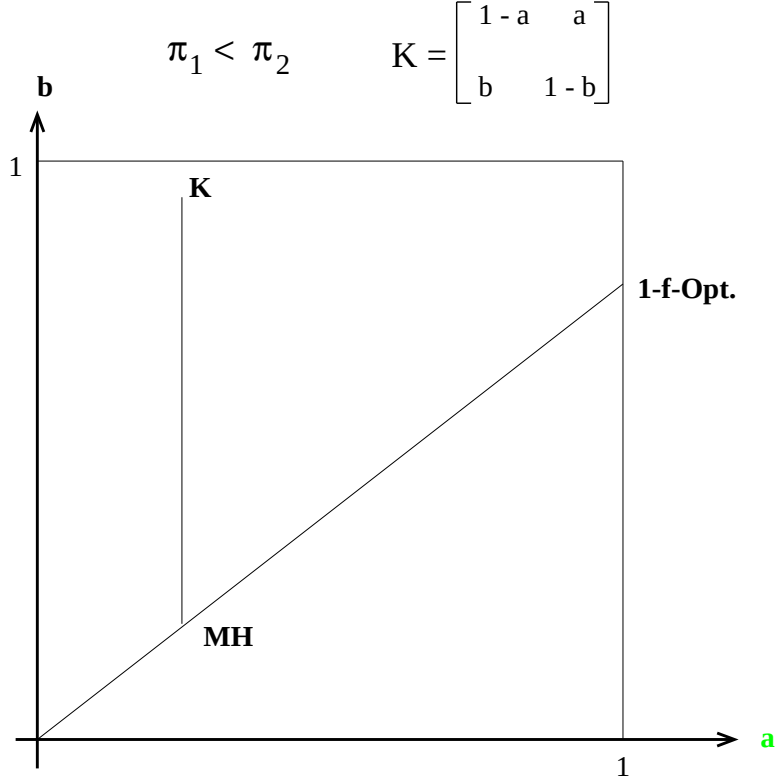


Figure 2: Geometrical interpretation

is also the optimal matrix with respect to uniform efficiency in this spacial case because every function is monotone on the state space considered.

4 Conclusions

A criterion that we have not yet mentioned, but that can be of practical use when choosing a Markov chain for MCMC purposes, is the ease of implementation of the underlying stochastic process. Once an algorithm is chosen we have to write computer code, using our favorite language, to run the simulation. This can be intensive work mostly due to the debugging process of the code. Thus a non-expert programmer might choose an easier to implement chain versus one that has slightly better performance. Regarding this point we note that the extra computational effort needed to insert a delaying rejection step into an existing MH or reversible jumps code is minimal.

All the orderings presented (except the convergence and the first degree efficiency ordering) are partial orderings. This means that they do not allow to compare all Markov chains that are either reversible or stationary with respect to the specific distribution of interest. We note that on the space of non-reversible Markov chains the covariance ordering, which can still be defined, is not a proper ordering anymore since the antisymmetry property fails to hold.

Acknowledgments

I am extremely grateful to Richard Tweedie for encouraging me to write the present manuscript and for insightful discussion on the topic treated here. I would also like to thank all my coauthors for helpful discussion and for sharing the intuition and the fun.

References

- Besag, J. and Green, P. J. (1993). Spatial statistics and Bayesian computation. *J. Roy. Statist. Soc. Ser. B* **55** 25–37.
- Billera, L. J. and Diaconis, P. (2001). A geometric interpretation of the metropolis algorithm. *Statistical Science*, this volume.
- Diaconis, P., Holmes, S., and Neal, R. M. (2000) Analysis of a non-reversible Markov chain sampler. *Annals of Applied Probability* **10** 726–752.
- Frigessi, A., Di Stefano, P., Hwang, A. and Sheu, A. (1993) Convergence rates of the Gibbs sampler, the Metropolis algorithm and other single-site updating dynamics. *J. R. Statist. Soc. B* **55** 205–219.
- Frigessi, A., Hwang, C. and Younes, L. (1992). Optimal spectral structure of reversible stochastic matrices, Monte carlo methods and the simulation of Markov random fields. *Ann. Appl. Probab.* **2** 610–628.
- Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82** 711–732.
- Green, P. J. and Mira, A. (1999). Delayed rejection in reversible jump Metropolis-Hastings. Technical report, Dep. of Mathematics, U. of Bristol, UK. *Biometrika*, under second revision. Paper available at <http://aim.unipv.it/~anto/delayed.ps>
- Liu, J. S. (1996a). Metropolized independent sampling. *Statistics and Computing* **6** 113–119.
- Liu, J. S. (1996b). Peskun theorem and a modified discrete-state Gibbs sampler. *Biometrika* **83** 681–682.
- Liu, J. S., Wong, W. H. and Kong, A. (1995). Correlation structure and convergence rate of the Gibbs sampler with various scans. *J. Roy. Statist. Soc. Ser. B* **57** 157–169.
- Mira, A. (2000). Efficiency increasing and stationarity preserving probability mass transfers for MCMC. Technical Report 3, University of Insubria, Dep. of Economics. Stat. and Prob. Letters, under second revision. Paper available at <http://aim.unipv.it/~anto/finite.ps>

- Mira, A. (2001). On Metropolis-Hastings algorithms with delayed rejection. *Metron*, to appear. Paper available at <http://aim.unipv.it/~anto/metron.ps>
- Mira, A. and Geyer, C. J. (1999). Ordering Monte Carlo Markov chains. Tech. Rep. 632, School of Statistics, University of Minnesota. Paper available at <http://aim.unipv.it/~anto/order.ps>
- Mira, A. and Geyer, C. J. (2000). On non-reversible Markov chains. *Fields Inst. Communic.: Monte Carlo Methods* **26** 93–108.
- Mira, A. and Tierney, L. (2001). Efficiency and convergence properties of slice samplers. *Scandinavian Journal of Statistics*, to appear. Paper available at <http://aim.unipv.it/~anto/aux.ps>
- Peskun, P. H. (1973). Optimum Monte Carlo sampling using Markov chains. *Biometrika* **60** 607–612.
- Roberts, G. O. (1996). Markov chain concepts related to sampling algorithms. In W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, editors, *Markov Chain Monte Carlo in Practice*. Chapman and Hall.
- Roberts, G. O. and Sahu, S. K. (1997) Updating schemes, covariance structure, blocking and parametrisation for the Gibbs sampler. *J. R. Statist. Soc. B* **59** 291–317.
- Roberts, G. O. and Sahu, S. K. (2001). Approximate pre-determined convergence properties of the Gibbs sampler. *J. Comput. and Graph. Statist.* **10**.
- Roberts, G. O. and Tweedie, R. L. (2001). *Understanding MCMC*. Springer Verlag.
- Sahu, S. K. and Roberts, G. O. (1999). On convergence of the EM algorithm and the Gibbs sampler. *Statistics and Computing* **9** 55–64.
- Tierney, L. (1998). A Note on Metropolis-Hastings kernels for general state spaces. *Annals of Applied Probability* **8** 1–9.
- Tierney, L. and Mira, A. (1999). Some adaptive Monte Carlo methods for Bayesian inference. *Statistics in Medicine* **18** 2507–2515.