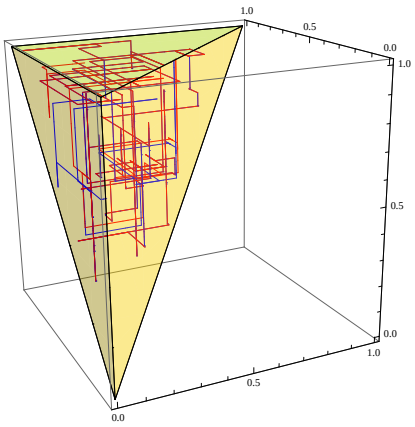
Ecole Normale Supérieure

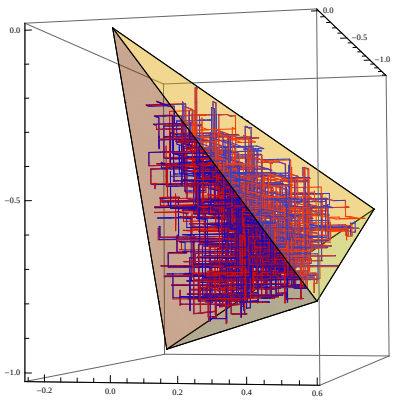
Internship report

Coupling of the heat-bath algorithm: from one-d hard spheres to high-dimensional polytopes

Author:
Sophie Cachot
ICFP Masters program

Supervisor:
Werner Krauth
CNRS, LPS





April 23, 2013

Contents

Abstract

The convergence of Markov chains is an important subject in statistics, computer science and computational physics. One of the main tools to prove convergence of Markov chains is *via* the coupling phenomenon, especially the "coupling from the past" approach [1]. Coupling is well established for a number of physically interesting models, such as the Ising model with heat-bath dynamics, as well as particle systems at low density, especially hard spheres [2, 3, 4]. There are many more applications in computer science, for example in graph theory.

Here, we study the heat-bath algorithm for hard spheres in one and two dimensions with a continuous configuration space. In one dimension, we show that this algorithm induces coupling of the entire configuration space: though the configurations never overlap exactly, they approach each other exponentially. For the one-dimensional problem, we determine the coupling time for a few small values of the number N of particles, and evaluate numerically the scaling of the coupling time with N.

Using a recently developed framework [5], we interpret the hard-sphere sampling problem in terms of coupling inside special polytopes. We observe rapid coupling for the heat-bath algorithm for the polytopes corresponding to one-dimensional sphere configurations, and seem to preserve rapid coupling for the two-dimensional case, also. However, the favorable convergence and coupling properties seem to disappear when we apply the heat-bath algorithm to more general polytopes.

1. Introduction

1.1 Markov chain Monte Carlo methods and mixing time

Markov chain Monte Carlo (MCMC) methods are a class of algorithms for sampling from probability distributions, sharing the following property: they build and run a Markov chain whose stationary distribution is the desired distribution. Thus, the state of the chain after a large number of iterations can be used as a sample of the target distribution. Since the quality of the sample improves with the number of steps, the main difficulty is to ascertain how many iterations are needed to converge to the equilibrium distribution, within an adequate error margin. This number of steps is the Markov chain mixing (or correlation) time, and is a measure of the efficiency of the algorithm.

The most common MCMC methods are random walks algorithms, moving in comparatively small steps around the equilibrium. A well-known example is the Metropolis algorithm, which generates moves using a jumping distribution and a method for rejecting proposed moves. A drawback of this algorithm is that the walk often covers ground already explored, so that it can take a long time to explore the whole configuration space. Devising MCMC algorithms with the same stationary distribution but exhibiting faster convergence is an active field of research.

In most cases, correlations decrease as $e^{-t/\tau}$, so the dependence on the starting point is negligible after a time $t \gg \tau$. However, the correlation time τ is usually unknown, and has to be determined through an empirical approach. This difficulty can sometimes be overcome using the coupling from the past approach [1], which enables one to determine the correlation time and obtain perfect samples. More generally, coupling arguments are amongst the most useful tools to prove convergence of Markov chains.

1.2 Heat-bath algorithm and coupling for the Ising model

Heat-bath algorithms form a general class of Markov chain Monte Carlo methods that at each step $t = 1, 2, ...$ thermalize a sub-system with respect to its local environment.

In the heat-bath algorithm for the Ising model (see, e.g., [6]), a randomly chosen spin is thermalized in the magnetic field produced by its neighbors. In the presence of the local field h, a spin points up and down with probability $\pi_h^+ = \frac{1}{1+e^{-2\beta h}}$ and $\pi_h^- = \frac{1}{1+e^{+2\beta h}}$, respectively. In order to sample this distribution, at each step, we pick a random site k, and a random number $\alpha = \text{random}(0, 1)$. We make the spin at site k point up if $\alpha < \pi_h^+$ and down otherwise, where h is the field at site k.

An interesting feature of this algorithm is that it preserves the natural half-order on configurations in the Ising model: a configuration $\{\sigma_1, ..., \sigma_N\}$ is smaller than another configuration $\{\sigma'_1, ..., \sigma'_N\}$ if $\sigma_k \leq \sigma'_k$ at each site k. With this definition, the two ground states $\boldsymbol{\sigma}_+$ and $\boldsymbol{\sigma}_-$ of the Ising model—all spins up or all spins down—are respectively larger and smaller than all configurations.

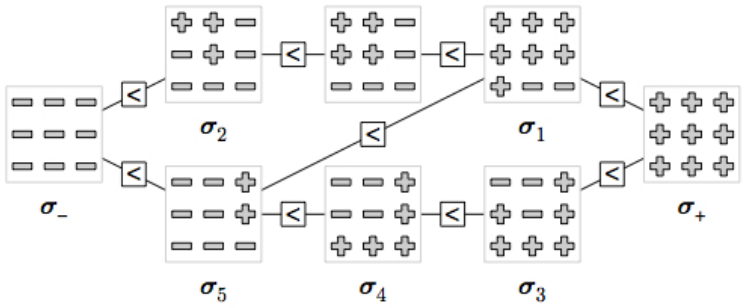


Figure 1.1: Half-order in the Ising model (taken from [6]): $\boldsymbol{\sigma}_-$ is smaller and $\boldsymbol{\sigma}_+$ is larger than all other configurations. Some configurations are unrelated (e.g., $\boldsymbol{\sigma}_2$ and $\boldsymbol{\sigma}_4$).

It is easy to check that the heat-bath algorithm preserves the half-order, provided we use the same random site k and the same random number α for all configurations. Thus, the coupling (or merging) of the two extremal configurations $\boldsymbol{\sigma}_+$ and $\boldsymbol{\sigma}_-$ indicates that all configurations have coupled.

This property leads to perfect sampling for the Ising model, using the coupling from the past framework (see [1]). If configurations $\boldsymbol{\sigma}_+$ and $\boldsymbol{\sigma}_-$ have coupled in n steps, we are certain that a sample obtained at time n from running the Markov chain with the same parameters k and α at each step is perfectly decorrelated from the initial condition. In practice, we implement this idea backward: at time n, we check whether the initial extremal configurations have merged. If so, we have obtained a perfect sample.

1.3 Heat-bath algorithm applied to hard spheres

The heat-bath algorithm can also be applied to hard-sphere models by thermalizing a randomly chosen sphere in the local environment of its neighbors, either directly in d-dimensional space [7], or in conveniently chosen submanifolds which simplify the calculations (see Fig. 1.2). We will follow the latter strategy, which is related to the hit-and-run framework [8].

The key difference with the Ising model is that the phase space is now continuous, and we cannot expect the heat-bath algorithm to couple exactly: two distinct initial configurations evolving under the algorithm will never overlap perfectly. However, they can approach each other, and this is what we will endeavor to characterize.

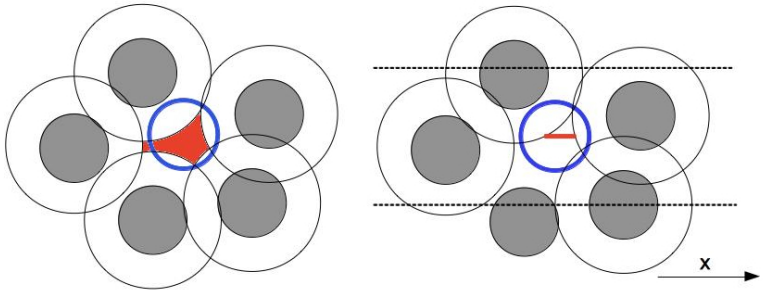


Figure 1.2: Heat-bath algorithms for two-dimensional hard spheres. *Left*: Algorithm of Watanabe et al. [7] — the updated position of the central sphere is uniformly sampled in its local d-dimensional neighborhood (*red area*). *Right*: The restriction to a lower-dimensional sub-manifold simplifies the sampling problem — the updated position is uniformly sampled on the accessible interval (*red line*).

The purpose of the present study is to analyze the heat-bath algorithm for hard spheres in dimensions one (Chapter 2) and two (Chapter 3), and to relate it to a more general sampling problem in polytopes (Chapter 4). In a nutshell, we show that although the heat-bath algorithm never "really" couples, configurations approach each other exponentially. We prove coupling times for a finite number $N = 2, ..., 7$ of spheres in one dimension, but must for the moment resort to numerics to obtain the scaling of the coupling time with N. For two-dimensional spheres, numerical results are still encouraging. However, the situation is more complex for general polytopes: the heat-bath algorithm does not seem to be a general coupling algorithm for the sampling in polytopes.

2. Heat-bath algorithm for one-d hard spheres

2.1 Definition, simplifications

For concreteness, we consider N one-dimensional hard spheres $0 \leq x_1 \leq x_2 \leq \cdots \leq x_N \leq L$ on a line of length L with fixed walls that we can represent by two additional, immobile spheres at $x_0 = 0$ and $x_{N+1} = L$. We can suppose without restrictions that the interval length L is equal to 1, and that the sphere radius σ is zero.

At each step $n = 0, 1, \ldots$ of the heat-bath algorithm for one-dimensional spheres, we do the following:

1. Sample a sphere k uniformly in $[1, N]$;
2. Thermalize its position with respect to its neighbors: $x_k \to x_{k-1} + \alpha(x_{k+1} - x_{k-1})$ with α uniform in $[0, 1]$.

2.2 Mixing (Correlation) time

The correlation time τ_{corr} of the above heat-bath algorithm should scale as $O(N^3)$. We only have a physicists' argument and a simulation (Fig. 2.1) to show this. Let us consider the mean position $m = \sum_k x_k/N$ which, in equilibrium, has fluctuations of order $\sigma(m) \sim \frac{1}{\sqrt{N}}$. On the other hand, one step of the algorithm moves a sphere on the scale of the inter-particle distance $\sim \frac{1}{N}$, inducing a change in m of order $\frac{1}{N^2}$. We remember that, for a random walk of step Δ, the walker's position after n steps has variance $\sigma^2 = n\Delta^2$. Here, we write $\frac{1}{\sqrt{N}} \sim \sqrt{\tau}_{\text{corr}} \frac{1}{N^2}$. Thus, we obtain $\tau_{\text{corr}} \sim N^3$. We stress that this intuitive argument is approximate, and does not take into account any less obvious dependence in N, such as $\log N$ factors.

2.3 Coupling - partial order

We may define a partial order of configurations x and y as follows:

$$x \leq y \Leftrightarrow x_i \leq y_i \ \forall i \in [1, N]. \tag{2.1}$$

This order is preserved under the heat-bath dynamics if we sample at time n the same values of k and of α for all configurations (see Fig. 2.2). There are two extremal configurations for this partial order: for any configuration x, $\hat{0} = (0, 0, \ldots, 0, 0) \leq x$ and $\hat{1} = (1, 1, \ldots, 1, 1) \geq x$.

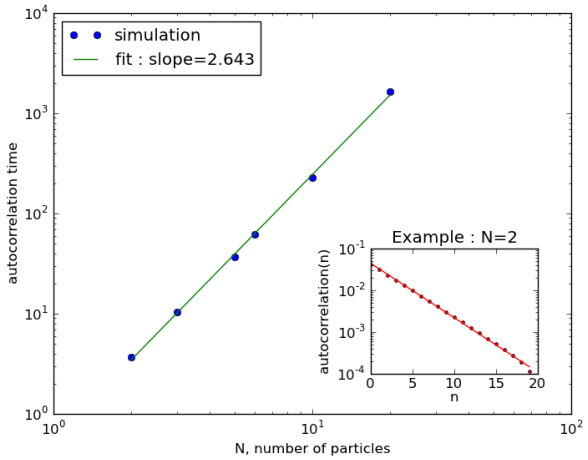


Figure 2.1: Autocorrelation time τ_{corr} of mean position m. *Inset*: Autocorrelation function for $N = 2$. An exponential fit for large times n yields the auto-correlation time.

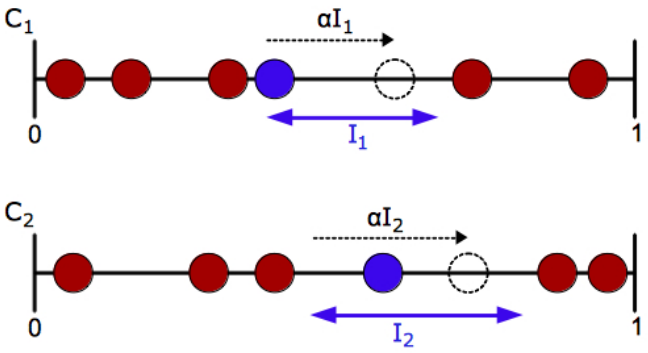


Figure 2.2: One step of the heat-bath coupling algorithm: we chose uniformly a particle $k \in [1, N]$ (*blue disk*) and a number $\alpha \in [0, 1]$, the same for both configurations. We move particle k at position αI in its available interval I. Note that $C_1 \leq C_2$ and that this order is conserved under the dynamics.

Two configurations x and y are said to be coupled if $\max_k |x_k - y_k| < b/N$, where b is arbitrary (in the simulations, we use $b = 10^{-10}$). To analyze the coupling behavior, we can restrict attention to the time evolution of the two extremal configurations $\hat{0}$ and $\hat{1}$: as with the Ising model, the coupling of these two configurations indicates that all configurations have coupled.

Simulations show exponential coupling for this algorithm. We will explain this behavior below, beginning with the case $N = 2$.

2.3.1 Coupling time - Binomial formula (two spheres)

We determine the evolution of the configurations $\hat{0}$ and $\hat{1}$ under simultaneous heat-bath dynamics. We call $\hat{0}(n)$ and $\hat{1}(n)$ the configurations after n updates from $\hat{0}$ and $\hat{1}$, respectively.

Let us write $y_k = x_k + \epsilon_k$ for all particles[1]. The heat-bath algorithm (see Section 2.1) has the following action on the ϵ:

$$(\epsilon_1, \epsilon_2) \to \begin{cases} (\alpha\epsilon_2, \epsilon_2) & \text{with prob. } 1/2 \\ (\epsilon_1, \alpha\epsilon_1) & \text{with prob. } 1/2 \end{cases}. \tag{2.2}$$

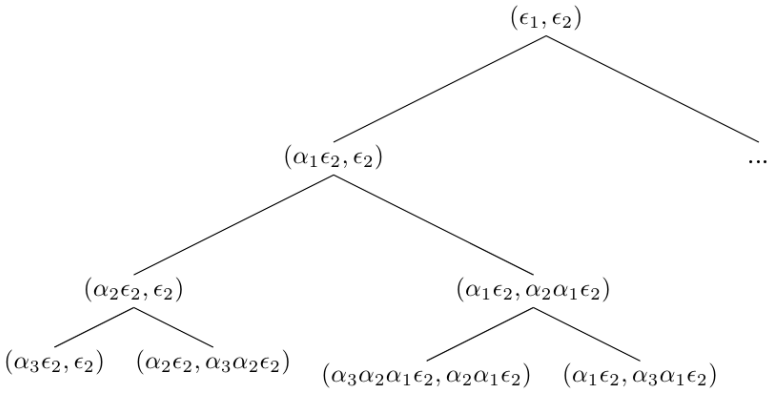


Figure 2.3: Coupling tree for the heat-bath algorithm for two spheres in one dimension (see Eq. (2.2), ϵ_1 and ϵ_2 describe the difference between two configurations, $\alpha_1, \alpha_2, \ldots$ are uniform random numbers in $[0, 1]$). The distance between the two configurations decreases with the number of switchbacks in the graph (changes left—right, or right—left). See Eq. (2.4) for the exact expected maximum distance for all times n.

Further iterations of the algorithm, starting from (ϵ_1, ϵ_2), build a "coupling tree". As readily seen in Fig. 2.3, the number of $\alpha's$ in $(y_1 - x_1, y_2 - x_2)$ increases with the number of switchbacks during the simulation: when we pick particle 1 after particle 2 (or vice-versa), we gain one α on both sides in $(y_1 - x_1, y_2 - x_2)$. On the other hand, if we pick particle 1 (or 2) twice in a row, one α is simply replaced by another.

Since at each step there is a probability $\frac{1}{2}$ of switching the particle to be moved, the number of switchbacks follows a binomial distribution. For $n \geq 2$, and for $K \in \{0, ..., n-1\}$, the probability of having K switchbacks up to time n is

$$P(K, n) = \binom{n-1}{K} \frac{1}{2^K} \frac{1}{2^{n-1-K}} = \binom{n-1}{K} \frac{1}{2^{n-1}} \tag{2.3}$$

[1]Beware: ϵ_k is not an infinitesimal, and the calculation is exact.

At time $n \geq 2$, the average maximum of $\hat{0}(n)$ and $\hat{1}(n)$ is given by

$$\begin{aligned}\langle \max(|x_1 - y_1|, |x_2 - y_2|) \rangle &= \sum_{K=0}^{N-1} P(K,n) \times \langle (\max(|x_1 - y_1|, |x_2 - y_2|)) \rangle_{K,n} \\ &= \sum_{K=0}^{N-1} \binom{n-1}{K} \frac{1}{2^{n-1}} \frac{1}{2^K} = \frac{1}{2^{n-1}} \left(1 + \frac{1}{2}\right)^{n-1} = \left(\frac{3}{4}\right)^{n-1} \end{aligned} \tag{2.4}$$

This exact formula demonstrates exponential coupling for the maximum of $|x_1 - y_1|$ and $|x_2 - y_2|$, with characteristic time $\tau_2 = -\frac{1}{\ln \frac{3}{4}} \approx 3,476$.

2.3.2 Coupling time - Transfer matrix ($N = 2 \dots 7$)

We have solved for the coupling timescale for $N = 2, 3, ..., 7$ using the transfer matrix of the coupling algorithm. Although we must discretize the transfer matrix, the second-largest eigenvalue seems to be independent of the discretization.

For $N = 2$, the recursion Eq. (2.2) can be discretized and represented through a transfer matrix. In order to discretize the configuration space of the (ϵ_1, ϵ_2) we call L the number of points on our discretized interval $]0, 1]$, and rescale by L, thus considering the lattice of points (L_1, L_2) with integer coordinates $1, 2, ..., L$.

We now associate to each discrete configuration (L_1, L_2) an index i, using the correspondence

$$(L_1, L_2) \longleftrightarrow i = (L_2 - 1)L + L_1 \ . \tag{2.5}$$

In the transfer matrix A, the element A_{ij} is the conditional probability of moving from index j to index i.

The non-zero elements of the first contribution to the matrix, corresponding to $(\epsilon_1, \epsilon_2) \to (\alpha\epsilon_2, \epsilon_2)$, are

$$A_1[(L_1, L_2), (L_1', L_2)] = \frac{1}{2L_2} \quad \text{for } 1 \leq L_1' \leq L_2, \tag{2.6}$$

whereas the second line of Eq. (2.2) yields

$$A_2[(L_1, L_2), (L_1, L_2')] = \frac{1}{2L_1} \quad \text{for } 1 \leq L_2' \leq L_1. \tag{2.7}$$

For $L = 2$, the transfer matrix is, using Eq. (2.5), Eq. (2.6), Eq. (2.7),

$$A = A_1 + A_2 = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{4} & 0 & \frac{1}{4} \\ 0 & 0 & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix} \quad (\text{for } N = 2, L = 2). \tag{2.8}$$

N	λ_2	τ_{coup}	τ_{coup}/N^3	Max. L
2	0.75	3.4761	0.4345	50
3	0.90237	9.7341	0.3605	15
4	0.95225	20.4402	0.3193	7
5	0.97321	36.8182	0.2945	5
6	0.98349	60.0856	0.2781	4
7	0.98913	95.4586	0.2783	3

Table 2.1: Coupling time-scale $\tau_{\text{coup}} = -1/\log\lambda_2$ for the heat-bath algorithm for one-dimensional hard spheres, obtained from the numerical diagonalization of the transfer matrix, analogous to Eq. (2.8). (NB: Transfer matrix A discretized on L^N points (see Eq. (2.8)) but λ_2 empirically independent of L.)

This matrix has a largest eigenvalue $\lambda_1 = 1$ and a corresponding eigenvector $\begin{bmatrix}1 & 0 & 0 & 0\end{bmatrix}$ that is concentrated on the smallest lattice point. Surprisingly, the second-largest eigenvalue $\lambda_2 = \frac{3}{4}$ is also independent of L. Since the convergence time is given by $\tau = -\frac{1}{\ln\lambda_2}$, we recover our previous result for τ_2.

Generalizing Eq. (2.2) for N particles, we obtain the following dynamics of the heat-bath algorithm on the parameters $\epsilon_1, ..., \epsilon_N$:

$$\epsilon_k \to (1-\alpha)\epsilon_{k-1} + \alpha\epsilon_{k+1} \quad \text{with prob. } 1/N \text{ for } k = 1 \dots N, \tag{2.9}$$

where we use $\epsilon_0 = \epsilon_{N+1} = 0$. We have diagonalized the discretized transfer matrix analogous to Eq. (2.8) for $N = 2, \dots, 7$, and computed λ_2 for small L, noting, as before, that it is independent of L. This yields a characteristic coupling time τ_{coup} which is nicely compatible with N^3 or $N^3 \log N$ scaling (see Table 2.1).

Our calculations in Table 2.1 are non-rigorous, because L is finite (while we are interested in λ_2 for $L \to \infty$) and because we compute λ_2 numerically. However, we observe that λ_2 is independent of L, something that we would like to understand. We would also like to be able to diagonalize the transfer matrix A for large values of N, in the limit $L \to \infty$.

2.3.3 Scaling of coupling time with N

We can study numerically the scaling of the characteristic coupling time τ with the number N of particles (Fig. 2.4). Additionally, we give the scaling of the real coupling time n_c for $b = 10^{-10}$, starting from the extremal initial conditions $\hat{0}$ and $\hat{1}$. In both cases, we observe a scaling compatible with N^3 or $N^3 \log N$ behavior for large values of N.

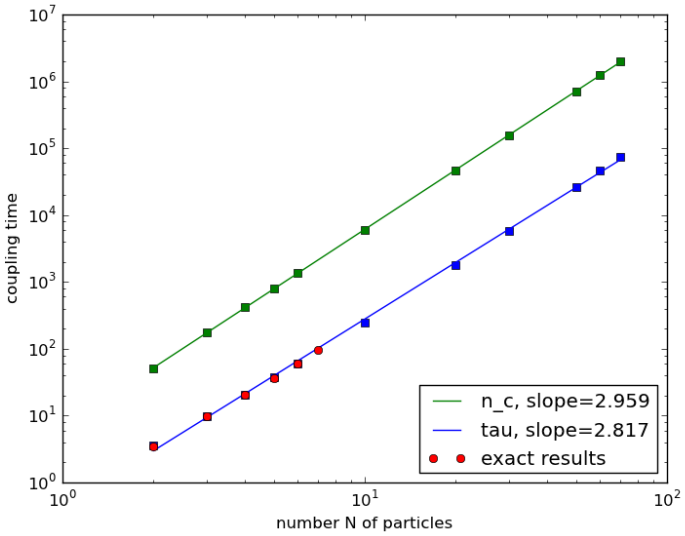


Figure 2.4: Scaling of the coupling times τ and n_c with N. The characteristic time τ is obtained both numerically and exactly (from the eigenvalue λ_2), while n_c is the average real coupling time obtained from simulations with $b = 10^{-10}$.

2.3.4 Understanding coupling

There are two ways to understand the coupling of the heat-bath algorithm. For once, coupling appears as a simple consequence of the preserved partial order with maximal elements [1]: both extremal configurations $\hat{0}(n)$ and $\hat{1}(n) \geq \hat{0}(n)$ undergo random walks. This is enough to prove that they must couple.

We can also use a convexity argument. Indeed, in Eq. (2.9), the expected value for $\langle \epsilon_k \rangle = (\epsilon_{k-1} + \epsilon_{k+1})/2$. This means that ϵ_k will end up on the straight line between its neighbors. With the boundary condition of $\epsilon_0 = \epsilon_{N+1}$, this shows (in a non-rigorous physicist's sense) that the long-time limit is the straight line with all the $\epsilon = 0$.

3. Two-d hard spheres - coupling time

3.1 Constraint graph

In a two-dimensional hard-sphere problem where spheres (here called disks) "slide" along an axis (e.g. along the x axis), the constraint graph—linking disks that may collide with each other—is fixed. It follows that the inequalities satisfied by the positions $(x_1, ..., x_N)$ of the non-overlapping disks remain unchanged when sliding the disks along x: the N-dimensional polytope they define is invariant [5]. As a consequence, the hard-disk sampling problem amounts to sampling inside a (special) polytope. We note that the one-dimensional heat-bath algorithm inside these polytopes is similar to the hit and run algorithm [8], except that we only consider moves along the coordinate axes, instead of sampling random directions inside the polytope.

We now consider the coupling problem for hard-disk configurations sharing the same constraint graph.

3.2 Simple example: $N = 3$ disks

We first consider $N = 3$ disks of radius σ, centered at (x_k, y_k) $(k = 1, 2, 3)$, with $x_1 \leq x_2 \leq x_3$ at time $n = 0$. Depending on the y_k, the situation can be qualitatively different from the one-dimensional case, and we may lose partial order. This happens when $y_2 \in [y_1 + \sigma, y_2 + 2\sigma[$ and $y_3 \in]y_1 - 2\sigma, y_1 - \sigma]$. In this situation, disk 1 can either collide with disk 2 or disk 3 at its right, while disk 2 and 3 cannot collide together (Fig. 3.1). This gives rise to more complicated steps of the algorithm, since the nearest right neighbor of disk 1 can be, for instance, disk 2 in configuration 1 and disk 3 in configuration 2.

The updating of ϵ_1 now depends on the positions of the disks, and we have to keep track of them in the transfer matrix. One state of the system is thus represented by $(\epsilon_1, \epsilon_2, \epsilon_3, x_1, x_2, x_3)$, and the discretized transfer matrix has L^6 elements [1] .

For small $L = 2, 3, 4$, we find that now the eigenvalues $\lambda = (1, ...)$ of the transfer matrix depend on the discretization step, except naturally for the highest eigenvalue $\lambda_1 = 1$ which is fixed by the conservation of probability. Thus, we obtain better approximation

[1] Since there are only 5 inequalities defining our situation, there should be a way to write a transfer matrix of size L^5.

of the coupling time with larger values of L: 6.7173 for $L = 2$, 6.5161 for $L = 3$ and 6.4355 for $L = 4$. This estimate agrees with our simulation results (see Fig. 3.1).

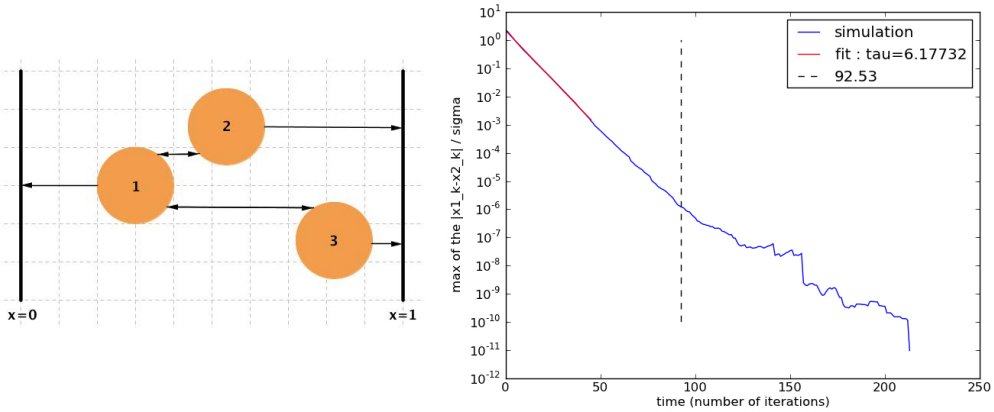


Figure 3.1: Coupling of a non-trivial situation with $N = 3$ spheres in two dimensions, averaged over 10000 runs. Half-order is lost, yet coupling remains exponential.

3.3 General case

We now assume periodic boundary conditions, and consider two configurations of N hard disks sharing the same constraint graph along the x-direction. Numerically, we see that the heat-bath algorithm along the x-axis still couples exponentially (at least for $N \lesssim 100$). Fig. 3.2 illustrates the coupling for two configurations of $N = 100$ particles.

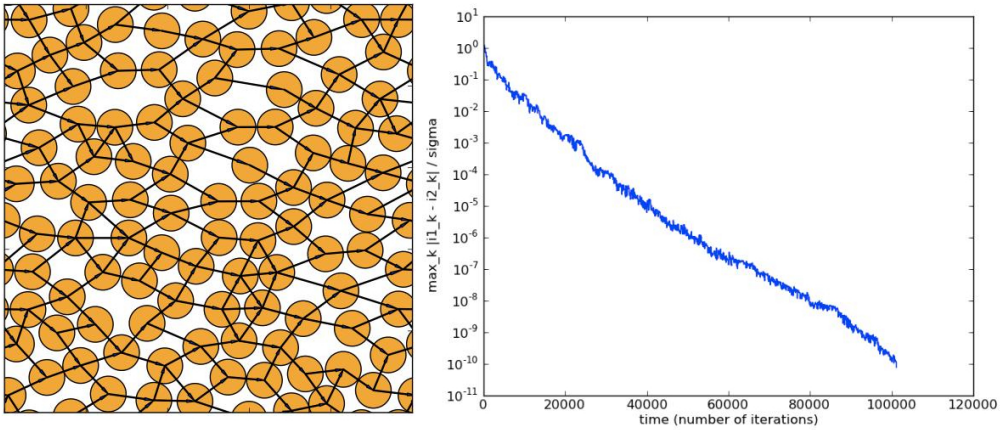


Figure 3.2: Coupling of two hard-disk configurations of $N = 100$ spheres in a two-dimensional periodic box, using the heat-bath algorithm along the x-axis. One of the configurations is depicted at the left, with its constraint graph along x: disks that may collide with each other are linked.

4. General polytopes

A general d-dimensional polytope can be specified by a set of n inequalities $Ax \leq b$ in $\mathbb{R}^d$, where A is a $n \times d$ matrix and b is a n-dimensional vector. In previous sections, we studied coupling inside specific polytopes (namely, those obtained from configurations of $1d$- or $2d$- hard spheres), whose defining inequalities were very simple in our chosen reference frame. We will now study the efficiency of the heat-bath algorithm in the context of generic polytopes, and consider the coupling as well as the convergence properties.

4.1 Rotated polytope of one-d hard spheres

Let us first examine a specific case: the polytope corresponding to one-d hard spheres, but described in a random orthonormal basis. In practice, we start from the defining inequalities $0 \leq x_1 \leq ... \leq x_N \leq 1$, where the dimension N of the polytope is the number of hard spheres, and generate a random orthogonal matrix M (see [6] ex. 1.15, p. 78). The new coordinates x' are given by $x = Mx'$, so the polytope is described by $(AM)x' \leq b$ in the new frame.

Depending on the coordinate system, we obtain a coupling which can be as quick as heat-bath coupling, or lose its exponential feature and show chaotic behavior. For instance, let us consider the three-dimensional case: if we rotate the frame so that the polytope no longer has edges along the axes (see Fig. 4.1), the mean coupling time (taken over runs with random initial configurations inside the polytope) can be multiplied by a factor of 15. The autocorrelation time also increases, but in a less striking way: for $d \leq 5$, it can be multiplied by a factor up to 3 depending on the coordinate system.

For $N \geq 4$, coupling do not always occur in our cutoff-time of $n_{cut} = 10^6$ iterations. In Table 4.1, we give averages of the real coupling time n_c (using $b = 10^{-10}$) taken over configurations that did couple before n_{cut}. Hence, the coupling time is all the more underestimated that the proportion of "non-coupling" runs is high. We see that the coupling time explodes with the dimension N when using a random orthonormal basis: the efficiency of the heat-bath algorithm for coupling inside the polytope of one-d hard spheres (Chapter 2) stems from a coordinate system which is well-fitted to the inequalities defining the polytope.

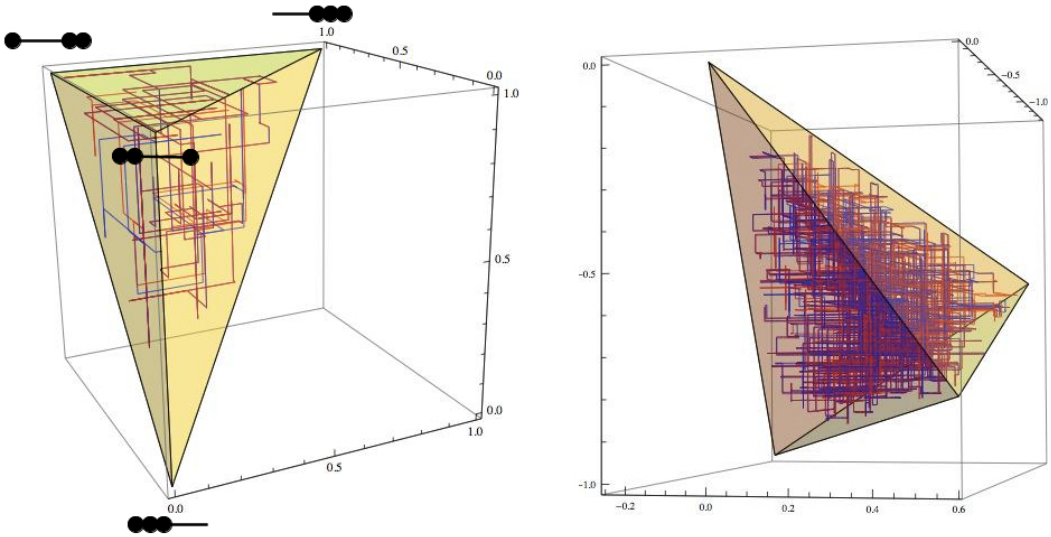


Figure 4.1: Coupling of two configurations inside the $3d$-polytope of one-dimensional hard spheres. *Left*: In the natural coordinates, $n_c \simeq 165$. Vertices are labelled by the corresponding configuration of three particles in the interval $[0, 1]$. *Right*: In a rotated frame, $n_c \simeq 2000$.

4.2 Random polytopes

We now build random polytopes in $\mathbb{R}$ by generating random inequalities, choosing the global sign of the coefficients of each inequality such that the origin lies inside the polytope. Coupling inside random polytopes yields results similar to those obtained previously, using a inadequate coordinate system for polytopes of one-d hard spheres. As before, the coupling behavior is almost always chaotic for polytopes with $d \geq 5$.

N	n_c standard basis	n_c (random orthonormal bases)	no coupling in 10^6 steps
2	47	$\times 2$	0%
3	165	$\times 3.3$	0%
4	383	$> \times 35$	1%
5	729	$\gg \times 170$	30%

Table 4.1: Coupling times n_c inside the polytope of one-d particles, using the standard coordinates or averaging over random orthonormal frames. The averages are taken over configurations which coupled in less than 10^6 iterations; hence, they may be underestimated.

5. Addendum: the article of D. Randall and P. Winkler

After this work was completed, we sent it for advice to the renowned mathematician P. Diaconis. He informed us that an article concerning the heat-bath algorithm for one-dimensional hard spheres [9] had been published in 2005, proving that the mixing time scales as $N^3 \log N$ with the number N of spheres. This article had escaped our notice, since it is not referenced in *Web of Science*, and can be found only on the webpage of the authors. We spotted a few inconsistencies, which we corrected and communicated to the authors; however, they do not invalidate the main conclusions. In the following, we relate the key points of the demonstration, which completes the study undertaken in Chapter 2.

The mathematical definition of the mixing time of a Markov chain X, whose stationary distribution is σ, is given by the minimum number t of steps such that the total variation distance

$$\|X(t) - \sigma\| = \max_A |P(X(t) \in A) - P(\sigma \in A)|$$

between $X(t)$ and σ is less than $1/4$, for any initial state of X.

The first part of the proof shows that the mixing time of the Markov chain described in Section 2.1 is $\mathcal{O}(N^3 \log N)$, using a coupling argument. Randall and Winkler consider the chain [1] $z(t) = \hat{1}(t) - \hat{0}(t)$, where the chains $\hat{1}(t)$ and $\hat{0}(t)$ are given by the evolution of the extremal configurations $\hat{1}$ and $\hat{0}$, respectively. We will write z for $z(t)$ and z' for $z(t+1)$.

The hard wall boundary conditions (see Section 2.1) read $z_0(t) = z_{N+1}(t) = 0\ \forall t$; they will play an important part below. The other key point is what we called the "convexity property" in Section 2.3.4: if sphere k is chosen at time t, then $\langle z'_k \rangle = (z_{k-1} + z_{k+1})/2$. Averaging over all choices of k, this gives $E[z'_k] = \frac{N-1}{N} z_k + \frac{1}{N}(z_{k-1} + z_{k+1})/2$.

Now, define

$$f(z) = \sum_{k=1}^{N} \sin\Big(\frac{k\pi}{N+1}\Big) z_k \ . \tag{5.1}$$

Using the two above properties, and the fact that [2] $\sin(a+b) + \sin(a-b) = 2 \sin a \cos b$,

[1] In Section 2.3, the chain z_k was denoted by ϵ_k.

[2] This justifies the choice of function f. We also use that $\sin\left(\frac{0}{N+1}\right) = \sin\left(\frac{(N+1)\pi}{N+1}\right) = 0$.

we can rewrite $E[f(z')]$ as a function of $f(z)$:

$$\begin{aligned} E[f(z')] &= \sum_{k=1}^{N} \sin\left(\frac{k\pi}{N+1}\right)\left(\frac{N-1}{N}z_k + \frac{1}{N}\frac{z_{k-1}+z_{k+1}}{2}\right) \\ &= \sum_{k=1}^{N}\left[\frac{N-1}{N}\sin\left(\frac{k\pi}{N+1}\right) + \frac{1}{2N}\left(\sin\left(\frac{(k+1)\pi}{N+1}\right) + \sin\left(\frac{(k-1)\pi}{N+1}\right)\right)\right] z_k \\ &= \sum_{k=1}^{N}\left[\frac{N-1}{N}\sin\left(\frac{k\pi}{N+1}\right) + \frac{1}{N}\cos\left(\frac{\pi}{N+1}\right)\sin\left(\frac{k\pi}{N+1}\right)\right] z_k \end{aligned}$$

so that

$$E[f(z')] = (1-\gamma_N) f(z) \tag{5.2}$$

where $\gamma_N = \frac{1}{N}\left(1 - \cos\frac{\pi}{N+1}\right)$.

Eq. (5.2) indicates that in average the function f decreases exponentially with the number of steps, with characteristic time $\tau_N = -1/\log(1-\gamma_N)$. This general formula coincides with the coupling times we obtained for $N = 2, ..., 7$ in Section 2.3.2, using the transfer matrix method. It also agrees with the result we found for $N = 2$ in Section 2.3.1, where we studied the decreasing with time of the average of $|\max z_k|$. The method used by Randall and Winkler is the same, but the function f they use is more adapted to the purpose, allowing them to find a formula valid for all N. More generally, functions with trigonometric weights like f can be used to prove mixing times for a wide class of problems [10].

In order to prove coupling in $\mathcal{O}(N^3 \log N)$ steps, the authors consider two phases of the coupling. In the first phase, of duration $t_1 \sim N^3 \log N$ steps, the spheres are moved according to the heat-bath algorithm, as explained in Fig. 2.2. This ensures that $\hat{0}(t)$ and $\hat{1}(t)$ become close: $z_k < 1.1\ N^{-4}$ (for N large enough) with probability at least 0.8 for all k. The second phase of the coupling is shorter, of duration $t_2 \sim N(\log N)^2$, and leads to exact coupling with probability almost 1. Hence, with probability almost 0.8, exact coupling takes place after $\sim N^3 \log N$ steps. Since the coupling time is necessarily larger than the mixing time, this proves the upper bound.

Randall and Winkler conclude the proof by showing that the mixing time is $\Omega(N^3 \log N)$. In order to prove this, they exhibit an event A such that $|P(X(t) \in A) - P(\sigma \in A)| > 1/4$ at $t_3 \sim N^3 \log N$. The event they consider is $A = [f(2x(t_3) - 1) > 0]$. For the stationary chain σ, it is obvious that $P(\sigma \in A) = 1/2$. On the other hand, using the chain $\hat{1}(t)$, it can be shown that $P(\hat{1}(t_3) \in A) > 3/4$. Thus, $\|X(t_3) - \sigma\| > 1/4$, proving that the chain has not mixed at time t_3.

6. Conclusion and outlook

The aim of the present study was to apply the heat-bath algorithm—well-known in the context of the Ising model—to hard spheres configurations, and evaluate its coupling time. The coupling time provides an upper bound for the mixing time, which is the crucial property of a Markov chain Monte Carlo algorithm. The main difficulty is that the phase space is now continuous, so that configurations can only "couple" approximately.

In one dimension, where a partial order can be defined as in the Ising model, we found that the heat-bath algorithm couples exponentially. Following an argument based on combinatorics, we evaluated the coupling time for $N = 2$ spheres. It would be interesting to know if this method can be generalized to any value of N. Using transfer matrices, we were also able to give the coupling times for $N = 2, ..., 7$ spheres. It is remarkable that our results coincide with those obtained by Randall and Winkler [9], given the discretization we used in order to define a transfer matrix. Besides, the results we found do not depend on the discretization step—something we have yet to understand. This work thus calls for further research, aiming to explain why the transfer matrix method leads to exact results.

In two dimensions, the half-order is lost—rendering an exact analysis arduous—, but our numerical results are encouraging: coupling is still exponential on individual runs when $N \lesssim 100$, for any density of hard disks. However, the scaling of the coupling time with N remains an open question. For two-dimensional hard spheres, it is known [11] that another algorithm (the "labelled displacement algorithm") induces rapid coupling for all N ($\tau_{coup}/N \sim a \log N + b$) up to a critical density $\eta_c \simeq 0.13$. It remains to be seen whether the heat-bath algorithm can do better.

Since the problem of hard-sphere sampling is related to sampling inside special polytopes, we investigated the efficiency of the heat-bath algorithm for more general polytopes: polytopes obtained by rotation of a one-d hard-sphere polytope, or random polytopes. In both cases, the coupling time increases drastically, indicating that the favorable properties of the heat-bath algorithm do not extend to generic polytopes.

A byproduct of this study lies in its didactic value: it can be used in teaching to illustrate in a simple way the notions of partial order, coupling, and rapid mixing in a continuous space.

References

[1] J. G. Propp, D. B. Wilson, *Random Struct. Algorithms* **9** 223-252 (1996).

[2] D. B. Wilson, *Random Struct. Algorithms* **16** 85-113 (2000).

[3] R. Kannan, M. W. Mahoney, R. Montenegro, in T. Ibaraki, N. Katoh, H. Ono (Eds.): Algorithms and Computation, 14th International Symposium, ISAAC 2003. Proceedings. Lecture Notes in Computer Science 2906, Springer (2003).

[4] C. Chanal, W. Krauth, *Phys. Rev. E* **81** 016705 (2010).

[5] S. Kapfer, W. Krauth, arXiv:1301.4901 to appear in *Journal of Physics A* (2013).

[6] W. Krauth, *Statistical Mechanics: Algorithms and Computations*, Oxford University Press (2006).

[7] H. Watanabe, S. Yukawa, M.A. Novotny, N. Ito, *Phys. Rev. E* **74** 026707 (2006).

[8] P. Diaconis, H. C. Andersen, *J. Soc. Francaise Stat.* **148** 5–28 (2007).

[9] D. Randall, P. Winkler, in C. Demetrescu, R. Sedgewick and R. Tamassia, eds, SIAM Press *Proc. 7th ALENEX & 2nd ANALCO 2005 (Vancouver BC)* 216-221 (2005).

[10] D. B. Wilson, *Annals of Appli. Probab.* **14** 274-325 (2004).

[11] E. P. Bernard, C. Chanal, W. Krauth, *EPL* **92** 60004 (2010).