

IMPROVED METROPOLIS–HASTINGS ALGORITHMS VIA LANDSCAPE MODIFICATION WITH APPLICATIONS TO SIMULATED ANNEALING AND THE CURIE–WEISS MODEL

MICHAEL C. H. CHOI ,* National University of Singapore

Abstract

In this paper, we propose new Metropolis–Hastings and simulated annealing algorithms on a finite state space via modifying the energy landscape. The core idea of landscape modification rests on introducing a parameter c , such that the landscape is modified once the algorithm is above this threshold parameter to encourage exploration, while the original landscape is utilized when the algorithm is below the threshold for exploitation purposes. We illustrate the power and benefits of landscape modification by investigating its effect on the classical Curie–Weiss model with Glauber dynamics and external magnetic field in the subcritical regime. This leads to a landscape-modified mean-field equation, and with appropriate choice of c the free energy landscape can be transformed from a double-well into a single-well landscape, while the location of the global minimum is preserved on the modified landscape. Consequently, running algorithms on the modified landscape can improve the convergence to the ground state in the Curie–Weiss model. In the setting of simulated annealing, we demonstrate that landscape modification can yield improved or even subexponential mean tunnelling time between global minima in the low-temperature regime by appropriate choice of c , and we give a convergence guarantee using an improved logarithmic cooling schedule with reduced critical height. We also discuss connections between landscape modification and other acceleration techniques, such as Catoni’s energy transformation algorithm, preconditioning, importance sampling, and quantum annealing. The technique developed in this paper is not limited to simulated annealing, but is broadly applicable to any difference-based discrete optimization algorithm by a change of landscape.

Keywords: Metropolis–Hastings; simulated annealing; spectral gap; Curie–Weiss; metastability; landscape modification; energy transformation

2020 Mathematics Subject Classification: Primary 60J27
Secondary 60J28

1. Introduction

Given a target Gibbs distribution with Hamiltonian function \mathcal{H} and temperature ϵ , the Metropolis–Hastings (MH) algorithm is a popular and important Markov chain Monte Carlo algorithm that has been applied to various sampling and optimization problems in a wide range of disciplines, including but not limited to Bayesian computation, statistical physics, and

Received 25 July 2022; accepted 19 July 2023.

* Postal address: Department of Statistics and Data Science, National University of Singapore, Singapore. Email address: mchchoi@nus.edu.sg

© The Author(s), 2023. Published by Cambridge University Press on behalf of Applied Probability Trust. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial licence (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use.

theoretical computer science. While there are many improved variants of the MH algorithm that have been investigated in the literature, this paper centres around a method that is known as landscape modification. This technique is particularly suitable in the context of stochastic minimization with respect to \mathcal{H} , either by running the MH algorithm at a low enough temperature ϵ or by driving the temperature ϵ to zero as in simulated annealing. The core idea of landscape modification relies on targeting a modified Hamiltonian function instead of \mathcal{H} , where the modification or the transformation is based upon two parameters, namely f and the threshold parameter c . On the part of landscape of \mathcal{H} that is below c , the original landscape is utilized, which allows for exploitation and concentration of the MH algorithm at low temperatures. On the other hand, on the region of the landscape of \mathcal{H} which is above c , the function f is applied to transform this part of the landscape to encourage and facilitate exploration of the chain. More precisely, the acceptance–rejection probability on this part is increased by the transformation, which consequently leads to a higher transition rate for the modified MH chain, which promotes exploration.

The advantage of landscape modification stems from the notion of critical height, which measures the difficulty of the landscape in a broad sense. With appropriate tuning of both f and c , the critical height is reduced, which leads to an improved simulated annealing algorithm or, for instance, reduced mean crossover time or relaxation time in the Curie–Weiss (CW) model, which we shall discuss in detail.

We now summarize and highlight the key achievements and contributions of this paper:

1. **We propose and analyse MH algorithms with landscape modification.** In Section 2, we first define a new MH algorithm using a modified Hamiltonian function. In general, the corresponding acceptance–rejection probability in the modified MH is of integral form, but we show that this integral can be readily calculated upon specializing to various choices of f , such as linear, quadratic, or square root functions. This is then followed by an example of landscape modification in the Ehrenfest urn with a linear Hamiltonian in Section 2.4, where we prove an upper bound on the spectral gap with polynomial dependence on the dimension, whereas the same technique yields an exponential dependence on the dimension for the classical MH. We provide a discussion on possible strategies for tuning f and c in Section 2.5. In the final subsection, that is, Section 2.6, we elaborate on the similarities and differences between landscape modification and other acceleration techniques in the Markov chain Monte Carlo literature, such as Catoni’s energy transformation algorithm [1, 2], preconditioning of the Hamiltonian, importance sampling, and quantum annealing [3].
2. **We obtain improved mean crossover time and relaxation time in the CW model with landscape modification.** In Section 3, we investigate the effect of landscape modification on the CW model. We first consider the CW model under a fixed external magnetic field in the subcritical regime, where the free energy landscape, as a function of the magnetization, has two local minima. Using a Glauber dynamics with landscape modification, we introduce a new mean-field equation, and with appropriate choice of c , we transform the free energy landscape from a double-well into a single-well landscape, while preserving the location of the global minimum on the modified landscape. As a result, running algorithms on the modified landscape can accelerate the convergence towards the ground state. We prove a subexponential mean tunnelling time in such a

setting and discuss related metastability results. Similar results are then extended to the random-field CW model.

3. **We obtain an improved simulated annealing algorithm with an improved logarithmic cooling schedule.** In Section 4, we consider the simulated annealing setting by driving the temperature down to zero. We define a clipped critical height c^* (that depends on the threshold parameter c) associated with the improved simulated annealing algorithm, and prove tight spectral gap asymptotics based on this parameter. This leads to similar asymptotic results concerning the total-variation mixing time and the mean tunnelling time in the low-temperature regime. Utilizing existing results concerning simulated annealing with time-dependent target function, we prove a convergence guarantee for the landscape-modified simulated annealing on a finite state space with an improved logarithmic cooling schedule. These theoretical results on improved convergence are corroborated by numerical experiments on a travelling salesman problem, in Section 4.1.

This paper can be seen as a sequel to an earlier work by the author [4]. The original motivation for landscape modification is from Fang *et al.* [5], who propose a variant of overdamped Langevin diffusion with state-dependent diffusion coefficient. Choi [4] applies this idea of state-dependent noise via landscape modification in the setting of kinetic simulated annealing and develops an improved kinetic simulated annealing algorithm with a convergence guarantee. In the present paper, we recognize that this idea of landscape modification can also be applied to the finite-state-space setting in MH and simulated annealing, and investigate the benefits and speed-ups that this technique can bring, in particular, to the analysis of the CW model and stochastic optimization. While the technique developed in Choi [4] can readily be applied to gradient-based continuous optimization algorithms, we emphasize that the landscape modification technique proposed in this paper can be analogously implemented in essentially all discrete optimization algorithms by a change of landscape and is not limited to simulated annealing or the CW model.

1.1. Notation

Throughout this paper, we adopt the following notation. For $x, y \in \mathbb{R}$, we let $x_+ = \max\{x, 0\}$ denote the non-negative part of x , and we write $x \wedge y = \min\{x, y\}$. For two functions $g_1, g_2 : \mathbb{R} \rightarrow \mathbb{R}$, we say that $g_1 = \mathcal{O}(g_2)$ if there exists a constant $C > 0$ such that for sufficiently large x , we have $|g_1(x)| \leq Cg_2(x)$. We write $g_1 = o(g_2)$ if $\lim_{x \rightarrow \infty} g_1(x)/g_2(x) = 0$, and we write $g_1 \sim g_2$ if $\lim_{x \rightarrow \infty} g_1(x)/g_2(x) = 1$. We say that $g_1(x)$ is a subexponential function if $\lim_{x \rightarrow \infty} \frac{1}{x} \log g_1(x) = 0$.

2. Metropolis–Hastings with landscape modification

Let \mathcal{X} be a finite state space, let $Q = (Q(x, y))_{x, y \in \mathcal{X}}$ be the transition matrix of a reversible proposal chain with respect to the probability measure $\mu = (\mu(x))_{x \in \mathcal{X}}$, and let $\mathcal{H} : \mathcal{X} \rightarrow \mathbb{R}$ be the target Hamiltonian function. Denote by $M^0 = (M^0(x, y))_{x, y \in \mathcal{X}}$ the infinitesimal generator of the continuized classical MH chain $X^0 = (X^0(t))_{t \geq 0}$, with proposal chain Q , and with target distribution being the Gibbs distribution $\pi^0(x) \propto e^{-\frac{1}{\epsilon} \mathcal{H}(x)} \mu(x)$ at temperature $\epsilon > 0$. Recall that its dynamics is given by

$$\begin{aligned}
 M^0(x, y) &= M_\epsilon^0(Q, \pi^0)(x, y) \\
 &:= \begin{cases} Q(x, y) \min \left\{ 1, e^{\frac{1}{\epsilon}(\mathcal{H}(x) - \mathcal{H}(y))} \right\} = Q(x, y)e^{-\frac{1}{\epsilon}(\mathcal{H}(y) - \mathcal{H}(x))_+} & \text{if } x \neq y, \\ -\sum_{z: z \neq x} M^0(x, z) & \text{if } x = y. \end{cases}
 \end{aligned}$$

We shall explain the superscript of 0 in both M^0 and π^0 in Definition 1 below.

Let us denote the ground-state energy level or the global minimum value of \mathcal{H} by $\mathcal{H}_{\min} := \min_{x \in \mathcal{X}} \mathcal{H}(x)$. Instead of directly targeting the Hamiltonian \mathcal{H} in the Gibbs distribution, we instead target the following modified or transformed function $\mathcal{H}_{\epsilon, c}^f$ at temperature ϵ :

$$\mathcal{H}_\epsilon(x) = \mathcal{H}_{\epsilon, c}^f(x) := \int_{\mathcal{H}_{\min}}^{\mathcal{H}(x)} \frac{1}{f((u - c)_+) + \epsilon} du, \tag{2.1}$$

where the function f and the parameter c are chosen to satisfy the following assumptions.

Assumption 1.

1. The function $f : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ is differentiable and non-decreasing, and it satisfies

$$f(0) = 0.$$

2. The parameter c satisfies $c \geq \mathcal{H}_{\min}$.

While it is impossible to calculate $\mathcal{H}_{\epsilon, c}^f$ without knowing \mathcal{H}_{\min} a priori, in an MH chain what matters is the difference of the energy function. For $x, y \in \mathcal{X}$, we see that

$$\mathcal{H}_{\epsilon, c}^f(y) - \mathcal{H}_{\epsilon, c}^f(x) = \int_{\mathcal{H}(x)}^{\mathcal{H}(y)} \frac{1}{f((u - c)_+) + \epsilon} du,$$

which does not depend on \mathcal{H}_{\min} . In the special case where we choose $f = 0$, the above equation reduces to $\mathcal{H}_{\epsilon, c}^0(y) - \mathcal{H}_{\epsilon, c}^0(x) = \frac{1}{\epsilon}(\mathcal{H}(y) - \mathcal{H}(x))$. On the other hand, in the case where $c < \mathcal{H}(x) < \mathcal{H}(y)$ and f is chosen such that $f(z) > 0$ whenever $z > 0$, we have

$$\mathcal{H}_{\epsilon, c}^f(y) - \mathcal{H}_{\epsilon, c}^f(x) \leq \frac{1}{f(\mathcal{H}(x) - c) + \epsilon} (\mathcal{H}(y) - \mathcal{H}(x)) < \frac{1}{\epsilon} (\mathcal{H}(y) - \mathcal{H}(x)).$$

Since f is assumed to be non-decreasing in Assumption 1, the greater the difference between $\mathcal{H}(x)$ and c , the smaller the upper bound is in the first inequality in the above equation; thus, the smaller we expect $\mathcal{H}_{\epsilon, c}^f(y) - \mathcal{H}_{\epsilon, c}^f(x)$ to be, and the higher the transition rate $Q(x, y) \exp \left\{ -(\mathcal{H}_{\epsilon, c}^f(y) - \mathcal{H}_{\epsilon, c}^f(x)) \right\}$ is. This is the sense in which the landscape is modified. Intuitively speaking, when the algorithm is above the threshold parameter c , the landscape is modified so that the transition rate is higher, to encourage exploration, while the original landscape is utilized for exploitation when the algorithm is below c .

To illustrate the effect of the proposed transformation, we plot the function

$$\mathcal{H}(x) = \cos(2x) + \frac{1}{2} \sin(x) + \frac{1}{3} \sin(10x)$$

as in Monmarché [6] and compare this landscape with that of $\mathcal{H}_{\epsilon, c}^f$ in Figure 1, where we take $\epsilon \in \{0.25, 0.5, 0.75, 1\}$, $c = -1.5$, and $f(z) = z$. With these choices of parameters, we compute

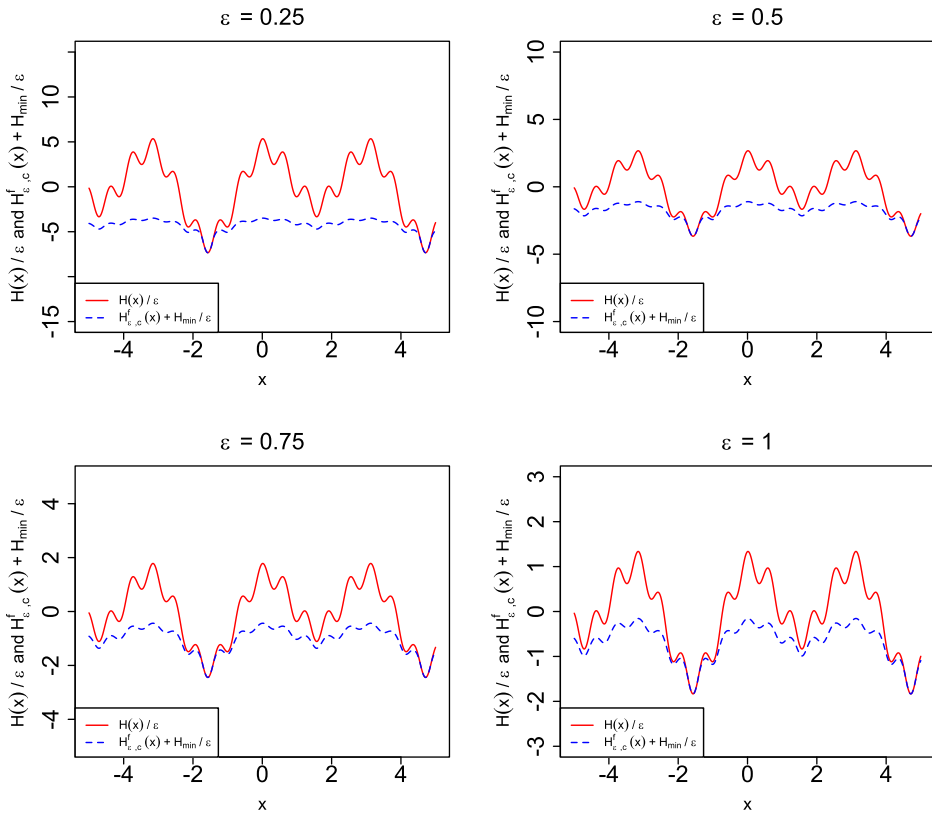


FIGURE 1. Landscape of $\frac{1}{\epsilon}\mathcal{H}$ and $\mathcal{H}_{\epsilon,c}^f + \frac{1}{\epsilon}\mathcal{H}_{\min}$, where $\mathcal{H}(x) = \cos(2x) + \frac{1}{2}\sin(x) + \frac{1}{3}\sin(10x)$, $\epsilon \in \{0.25, 0.5, 0.75, 1\}$, $c = -1.5$, and $f(z) = z$.

that

$$\mathcal{H}_{\epsilon,c}^f(x) = \frac{1}{\epsilon}(\min\{c, \mathcal{H}(x)\} - \min \mathcal{H}) + \ln\left(1 + \frac{1}{\epsilon}(\mathcal{H}(x) - c)_+\right).$$

In view of the above equation, we plot and compare $\frac{1}{\epsilon}\mathcal{H}(x)$ and $\mathcal{H}_{\epsilon,c}^f(x) + \frac{1}{\epsilon}\mathcal{H}_{\min}$ in Figure 1 on the domain $\mathcal{X} = \{-5 + \frac{k}{1000}10; k = 1, 2, \dots, 1000\}$. The shift of $\frac{1}{\epsilon}\mathcal{H}_{\min}$ is necessary to make these two landscapes match exactly in the region where $\{x; \mathcal{H}(x) \leq c\}$, so that they are on the same scale. When $x \in (-4, -2)$ in Figure 1, we can see that the gradient of $\mathcal{H}_{\epsilon,c}^f$ is smaller than that of $\frac{1}{\epsilon}\mathcal{H}$; thus it is easier to climb up the hill in this region (i.e. there is a higher transition rate). From the plot we also note that both $\frac{1}{\epsilon}\mathcal{H}$ (the solid red curve) and $\mathcal{H}_{\epsilon,c}^f$ (the dashed blue curve) have exactly the same set of stationary points.

Keeping in mind the ideas and notation above, we are now ready to introduce the MH chain with landscape modification.

Definition 1. (*Metropolis–Hastings with landscape modification.*) Let \mathcal{H} be the target Hamiltonian function, and let $\mathcal{H}_{\epsilon,c}^f$ be the landscape-modified function at temperature ϵ introduced in (2.1), where f and c satisfy Assumption 1. The continuized Metropolis–Hastings

chain with landscape modification $X_{\epsilon,c}^f = (X_{\epsilon,c}^f(t))_{t \geq 0}$ has target distribution $\pi^f(x) = \pi_{\epsilon,c}^f(x) \propto e^{-\mathcal{H}_{\epsilon,c}^f(x)} \mu(x)$ and proposal chain Q , and its infinitesimal generator $M^f = (M^f(x, y))_{x,y \in \mathcal{X}}$ is given by

$$M^f(x, y) = M_{\epsilon,c}^f(Q, \pi^f)(x, y) := \begin{cases} Q(x, y)e^{-(\mathcal{H}_{\epsilon,c}^f(y) - \mathcal{H}_{\epsilon,c}^f(x))_+} & \text{if } x \neq y, \\ -\sum_{z: z \neq x} M^f(x, z) & \text{if } x = y. \end{cases}$$

Note that when $f = 0$, the above dynamics reduces to the classical MH X^0 .

We now fix some notation and recall some important concepts and results from the literature on Markov chains. We endow the Hilbert space $\ell^2(\pi_{\epsilon,c}^f)$ with the usual inner product weighted by the invariant measure $\pi_{\epsilon,c}^f$: for $g_1, g_2 \in \ell^2(\pi_{\epsilon,c}^f)$,

$$\langle g_1, g_2 \rangle_{\pi_{\epsilon,c}^f} := \sum_{x \in \mathcal{X}} g_1(x)g_2(x)\pi_{\epsilon,c}^f(x),$$

and for $p > 1$ we denote the ℓ^p norm by $\|\cdot\|_{\ell^p(\pi_{\epsilon,c}^f)}$. We write $\lambda_2(-M_{\epsilon,c}^f)$ for the spectral gap of $M_{\epsilon,c}^f$; that is,

$$\lambda_2(-M_{\epsilon,c}^f) := \inf_{l \in \ell^2(\pi_{\epsilon,c}^f) : \pi_{\epsilon,c}^f(l) = 0} \frac{\langle -M_{\epsilon,c}^f l, l \rangle_{\pi_{\epsilon,c}^f}}{\langle l, l \rangle_{\pi_{\epsilon,c}^f}}. \tag{2.2}$$

Analogously, we denote by $\lambda_2(-M_{\epsilon}^0)$ the spectral gap of M_{ϵ}^0 .

In the upcoming sections, we shall investigate and compare the total-variation mixing time between the MH chains with and without landscape modification. For any probability measures ν_1, ν_2 with support on \mathcal{X} , the total-variation distance between ν_1 and ν_2 is

$$\|\nu_1 - \nu_2\|_{TV} := \sup_{A \subset \mathcal{X}} |\nu_1(A) - \nu_2(A)| = \frac{1}{2} \sum_{x \in \mathcal{X}} |\nu_1(x) - \nu_2(x)|.$$

The worst-case total-variation mixing time is defined to be

$$t_{mix}(M_{\epsilon,c}^f, 1/4) := \inf \left\{ t; \sup_x \|P_t^f(x, \cdot) - \pi_{\epsilon,c}^f\|_{TV} < 1/4 \right\},$$

where $(P_t^f = e^{M_{\epsilon,c}^f t})_{t \geq 0}$ is the transition semigroup of $X_{\epsilon,c}^f$.

Besides the mixing time, we will also be interested in various hitting times of the MH chain. These variables naturally appear when we discuss metastability results in the CW model in Section 3 or in discrete simulated annealing in Section 4. For any $A \subset \mathcal{X}$, we write $\tau_A^f := \inf \{t \geq 0; X_{\epsilon,c}^f(t) \in A\}$, and the usual convention that $\inf \emptyset = \infty$ applies. Similarly, we define τ_A^0 to be the first hitting time of the set A for the MH chain X^0 . When $A = \{x\}$, we shall simply write $\tau_x^f = \tau_{\{x\}}^f$ (resp. $\tau_x^0 = \tau_{\{x\}}^0$). Also, we denote by $\mathbb{E}_x(\cdot)$ (resp. $\mathbb{P}_x(\cdot)$) the mathematical expectation (resp. probability) of the Markov chain with initial state at $x \in \mathcal{X}$.

One crucial notion that quantifies the possible benefits of landscape modification is the concept of critical height, which in a broad sense measures the difficulty of the landscape. We

now recall this classical notion, which originates from the literature on simulated annealing and metastability. A path from x to y is any sequence of points $x_0 = x, x_1, x_2, \dots, x_n = y$ such that $Q(x_{i-1}, x_i) > 0$ for $i = 1, 2, \dots, n$. For any $x \neq y$, such a path exists as the proposal chain Q is irreducible. We write $\chi^{x,y}$ for the set of paths from x to y , and elements of $\chi^{x,y}$ are denoted by $\gamma = (\gamma_i)_{i=0}^n$. Given a target function \mathcal{U} defined on \mathcal{X} , the highest value of \mathcal{U} along a path $\gamma \in \chi^{x,y}$, known as the elevation, is defined to be

$$\text{Elev}(\mathcal{U}, \gamma) := \max\{\mathcal{U}(\gamma_i); \gamma_i \in \gamma\},$$

and the lowest possible highest elevation along paths from x to y is

$$G(x, y) = G(\mathcal{U}, x, y) := \min\{\text{Elev}(\mathcal{U}, \gamma); \gamma \in \chi^{x,y}\}. \tag{2.3}$$

The associated critical height of \mathcal{U} is then defined to be

$$L(\mathcal{U}) := \max_{x,y \in \mathcal{X}} \{G(x, y) - \mathcal{U}(x) - \mathcal{U}(y)\} + \min \mathcal{U}.$$

Another related notion is the clipped critical height c^* , which is defined to be

$$c^*(\mathcal{U}, c) := \max_{x,y \in \mathcal{X}} \{(G(\mathcal{U}, x, y) \wedge c) - (\mathcal{U}(x) \wedge c) - (\mathcal{U}(y) \wedge c)\} + \min \mathcal{U}.$$

One can understand c^* as if we are optimizing with respect to the function $\mathcal{U} \wedge c$. In Section 3, we shall consider \mathcal{U} to be the free energy with and without landscape modification that arises in the CW model, while in Section 4, when we investigate an improved simulated annealing algorithm, we take \mathcal{U} to be either \mathcal{H} or $\mathcal{H}_{\epsilon,c}^f$. We refer readers to Figure 4 in Section 4, where we offer a visual illustration of the notion of critical height.

To simulate $X_{\epsilon,c}^f$ practically, we would need to evaluate the acceptance–rejection probability, which amounts to the following integration:

$$\begin{aligned} & \exp\left(-(\mathcal{H}_{\epsilon,c}^f(y) - \mathcal{H}_{\epsilon,c}^f(x))_+\right) \\ &= \begin{cases} 1 & \text{if } \mathcal{H}(y) \leq \mathcal{H}(x), \\ \exp\left(\frac{1}{\epsilon}(\mathcal{H}(y) - \mathcal{H}(x))\right) & \text{if } c \geq \mathcal{H}(y) > \mathcal{H}(x), \\ \exp\left(\frac{1}{\epsilon}(c - \mathcal{H}(x)) - \int_c^{\mathcal{H}(y)} \frac{1}{f(u-c) + \epsilon} du\right) & \text{if } \mathcal{H}(y) > c \geq \mathcal{H}(x), \\ \exp\left(-\int_{\mathcal{H}(x)}^{\mathcal{H}(y)} \frac{1}{f(u-c) + \epsilon} du\right) & \text{if } \mathcal{H}(y) > \mathcal{H}(x) > c. \end{cases} \tag{2.4} \end{aligned}$$

In the following three subsections, we evaluate the above integrals (2.4) with f chosen to be a linear, quadratic, or cubic function, respectively.

2.1. Linear f : Metropolis–Hastings with logarithmic Hamiltonian and Catoni’s energy transformation algorithm

In this subsection, we specialize to $f(u) = u$. It turns out we can understand the landscape modification as if the Hamiltonian is on a logarithmic scale whenever $\mathcal{H}(x) > c$.

For $x, y \in \{\mathcal{H}(y) > \mathcal{H}(x) \geq c\}$, since

$$\int_{\mathcal{H}(x)}^{\mathcal{H}(y)} \frac{1}{u - c + \epsilon} du = \ln \left(\frac{\mathcal{H}(y) - c + \epsilon}{\mathcal{H}(x) - c + \epsilon} \right),$$

substituting the expression into (2.4) gives

$$\exp(-(\mathcal{H}_{\epsilon,c}^f(y) - \mathcal{H}_{\epsilon,c}^f(x))_+) = \begin{cases} 1 & \text{if } \mathcal{H}(y) \leq \mathcal{H}(x), \\ \exp\left(-\frac{1}{\epsilon}(\mathcal{H}(y) - \mathcal{H}(x))\right) & \text{if } c \geq \mathcal{H}(y) > \mathcal{H}(x), \\ \exp\left(-\frac{1}{\epsilon}(c - \mathcal{H}(x))\right) \frac{\epsilon}{\mathcal{H}(y) - c + \epsilon} & \text{if } \mathcal{H}(y) > c \geq \mathcal{H}(x), \\ \frac{\mathcal{H}(x) - c + \epsilon}{\mathcal{H}(y) - c + \epsilon} & \text{if } \mathcal{H}(y) > \mathcal{H}(x) > c. \end{cases}$$

This resulting dynamics M^f coincides with the energy transformation method introduced by Catoni [1, 2] on $\{\mathcal{H}(y) > \mathcal{H}(x) > c\}$, which is based on a logarithmic Hamiltonian. We refer readers to Section 2.6 for a more detailed account of the connection between landscape modification and Catoni’s energy transformation.

2.2. Quadratic f : Metropolis–Hastings with arctan Hamiltonian

In this subsection, we take $f(u) = u^2$. In this case, the effect of the landscape modification gives an inverse-tangent-transformed Hamiltonian whenever $\mathcal{H}(x) > c$.

For $x, y \in \{\mathcal{H}(y) > \mathcal{H}(x) \geq c\}$, using the inverse-tangent difference formula we obtain

$$\begin{aligned} \int_{\mathcal{H}(x)}^{\mathcal{H}(y)} \frac{1}{(u - c)^2 + \epsilon} du &= \sqrt{\frac{1}{\epsilon}} \left(\arctan\left(\sqrt{\frac{1}{\epsilon}}(\mathcal{H}(y) - c)\right) - \arctan\left(\sqrt{\frac{1}{\epsilon}}(\mathcal{H}(x) - c)\right) \right) \\ &= \sqrt{\frac{1}{\epsilon}} \arctan\left(\frac{\sqrt{\frac{1}{\epsilon}}(\mathcal{H}(y) - \mathcal{H}(x))}{1 + \frac{1}{\epsilon}(\mathcal{H}(y) - c)(\mathcal{H}(x) - c)}\right), \end{aligned}$$

and substituting the above expression into (2.4) gives

$$\exp(-(\mathcal{H}_{\epsilon,c}^f(y) - \mathcal{H}_{\epsilon,c}^f(x))_+) = \begin{cases} 1 & \text{if } \mathcal{H}(y) \leq \mathcal{H}(x), \\ \exp\left(\frac{1}{\epsilon}(\mathcal{H}(y) - \mathcal{H}(x))\right) & \text{if } c \geq \mathcal{H}(y) > \mathcal{H}(x), \\ \exp\left(\frac{1}{\epsilon}(c - \mathcal{H}(x)) - \sqrt{\frac{1}{\epsilon}} \arctan\left(\sqrt{\frac{1}{\epsilon}}(\mathcal{H}(y) - c)\right)\right) & \text{if } \mathcal{H}(y) > c \geq \mathcal{H}(x), \\ \exp\left(\sqrt{\frac{1}{\epsilon}} \left(\arctan\left(\sqrt{\frac{1}{\epsilon}}(\mathcal{H}(x) - c)\right) - \arctan\left(\sqrt{\frac{1}{\epsilon}}(\mathcal{H}(y) - c)\right) \right)\right) & \text{if } \mathcal{H}(y) > \mathcal{H}(x) > c. \end{cases}$$

2.3. Square root f : Metropolis–Hastings with sum of square root and logarithmic Hamiltonian

In the final example, we let $f(z) = \sqrt{z}$ for $z \geq 0$. For $x, y \in \{\mathcal{H}(y) > \mathcal{H}(x) \geq c\}$, consider the integral

$$\int_{\mathcal{H}(x)}^{\mathcal{H}(y)} \frac{1}{\sqrt{u-c} + \epsilon} du = 2 \left(\left(\sqrt{\mathcal{H}(y)-c} - \sqrt{\mathcal{H}(x)-c} \right) - \epsilon \ln \left(\frac{\sqrt{\mathcal{H}(y)-c} + \epsilon}{\sqrt{\mathcal{H}(x)-c} + \epsilon} \right) \right).$$

Substituting the expression into (2.4) gives

$$\begin{aligned} & \exp\left(-(\mathcal{H}_{\epsilon,c}^f(y) - \mathcal{H}_{\epsilon,c}^f(x))_+\right) \\ &= \begin{cases} 1 & \text{if } \mathcal{H}(y) \leq \mathcal{H}(x), \\ \exp\left(-\frac{1}{\epsilon}(\mathcal{H}(y) - \mathcal{H}(x))\right) & \text{if } c \geq \mathcal{H}(y) > \mathcal{H}(x), \\ \exp\left(-\frac{1}{\epsilon}(c - \mathcal{H}(x)) - 2\sqrt{\mathcal{H}(y)-c}\right) \left(\frac{\sqrt{\mathcal{H}(y)-c} + \epsilon}{\epsilon}\right)^{2\epsilon} & \text{if } \mathcal{H}(y) > c \geq \mathcal{H}(x), \\ \exp(2\sqrt{\mathcal{H}(x)-c} - 2\sqrt{\mathcal{H}(y)-c}) \left(\frac{\sqrt{\mathcal{H}(y)-c} + \epsilon}{\sqrt{\mathcal{H}(x)-c} + \epsilon}\right)^{2\epsilon} & \text{if } \mathcal{H}(y) > \mathcal{H}(x) > c. \end{cases} \end{aligned}$$

Remark 1. (*Use of landscape modification for sampling.*) This paper focuses on investigating the acceleration effect of landscape modification in stochastic optimization and simulated annealing. However, the technique of landscape modification can also be applied to sampling from multimodal distributions. The paper Zhang and Choi [7] analyses the use of landscape modification for sampling. We now briefly describe the setting therein. Suppose that we are interested in sampling from a multimodal distribution that we denote by $\nu(x) \propto e^{-\mathcal{H}(x)}$. Applying the idea of landscape modification, we then construct an MH chain with the transformed Hamiltonian function

$$\mathcal{H}_{1,c,\alpha}^f(x) = \int_{\mathcal{H}_{\min}}^{\mathcal{H}(x)} \frac{1}{\alpha f((u-c)_+) + 1} du, \tag{2.5}$$

so that its stationary distribution is given by $\nu_{1,c,\alpha}^f(x) \propto e^{-\mathcal{H}_{1,c,\alpha}^f(x)}$. We note that the parameter $\alpha \geq 0$, introduced in (2.5), controls the bias between the distribution ν and its landscape-modified counterpart $\nu_{1,c,\alpha}^f$. If we anneal this parameter by sending $\alpha_t \rightarrow 0$ as $t \rightarrow \infty$, then ν_{1,c,α_t}^f converges weakly to the target distribution ν . As a result, we are able to construct a non-homogeneous MH chain that converges to ν in the long run while enjoying the benefits of landscape modification.

2.4. A Metropolized Ehrenfest urn with landscape modification

In this section, we discuss a Metropolized Ehrenfest urn model; our exposition closely follows that of Deuschel and Mazza ([8], Section 5.1.2). Let us first briefly fix the setting. We consider the state space $\mathcal{X} = \{0, 1, \dots, d\}$ with $d \in \mathbb{N}$ and take a linear Hamiltonian $\mathcal{H}(x) = x$, where $\mathcal{H}_{\min} = 0$. The proposal birth–death chain has generator Q given by $Q(x, x+1) = 1 - x/d$, $Q(x, x-1) = x/d$, $Q(x, x) = -1$ and zero otherwise, and we note that the stationary measure of Q is $\mu(x) \propto 2^{-d} \binom{d}{x}$. With these choices, the classical Metropolized dynamics is

$M_\epsilon^0(x, x + 1) = Q(x, x + 1)e^{-\frac{1}{\epsilon}}$, $M_\epsilon^0(x, x - 1) = Q(x, x - 1)$, and $\pi^0(x) \propto e^{-\frac{1}{\epsilon}x}\mu(x)$. It is shown in Deuschel and Mazza ([8], Equation (5.1.1)) that

$$\lambda_2(-M_\epsilon^0) \leq \lambda_2(-Q) \frac{1}{\sum_{x=0}^d e^{-\frac{1}{\epsilon}x}\mu(x)} = \frac{d}{2} \frac{2^d}{(1 + e^{-\frac{1}{\epsilon}})^d}. \tag{2.6}$$

At a fixed temperature ϵ , we note that the upper bound in (2.6) is exponential in d .

Now, we consider the landscape-modified MH with $f(z) = z$ and $c = 1$. With these parameters, we compute

$$\mathcal{H}(x) = \int_0^x \frac{1}{(u - 1)_+ + \epsilon} du = \frac{1}{\epsilon} + \ln\left(\frac{(x - 1)_+ + \epsilon}{\epsilon}\right).$$

Using Deuschel and Mazza ([8], Equation (5.1.1)) leads to

$$\lambda_2(-M_{\epsilon,c=1}^f) \leq \lambda_2(-Q) \frac{1}{\sum_{x=0}^d e^{-\mathcal{H}(x)}\mu(x)} = \frac{d}{2} \frac{2^d e^{\frac{1}{\epsilon}}}{\frac{\epsilon}{\lfloor d/2 \rfloor - 1 + \epsilon} \binom{d}{\lfloor d/2 \rfloor}} \sim d^3 \frac{1}{\epsilon} e^{\frac{1}{\epsilon}}, \tag{2.7}$$

where we use Stirling’s formula, which gives, for large enough d ,

$$\binom{d}{\lfloor d/2 \rfloor} \sim \frac{2^d}{\sqrt{\pi \lfloor d/2 \rfloor}}.$$

As a result, the upper bound in (2.7) gives a polynomial dependency on d , at the tradeoff of an exponential dependency on $\frac{1}{\epsilon}$. In retrospect this result is not surprising, since we are working with a logarithmic Hamiltonian and hence the asymptotics is in the polynomial of d instead of 2^d . This section serves as a warm-up example to prepare us for discussing the more complicated CW model in Section 3.

2.5. Tuning strategies for f and c

In this section, we discuss tuning strategies for f and the threshold parameter c in the context of using the MH chain with landscape modification for stochastic optimization.

First, the function f controls how the landscape is transformed above the threshold parameter c . For example, taking a linear $f(z) = z$ gives a logarithmic transformation, while taking a quadratic $f(z) = z^2$ yields an arctan transformation, as shown in Sections 2.1 and 2.2. In general, we recommend using $f(z) = z^2$, since the arctan transformation is uniformly bounded above by $\pi/2$, which facilitates exploration on the part of the landscape above c by giving a higher transition rate compared with the choice of $f(z) = z$. However, we suspect that in numerical investigations, and depending on the target Hamiltonian \mathcal{H} , there are possibilities of using a linear f to yield improved convergence towards the global minimum.

Second, as we shall see in Section 4, the threshold parameter c controls the clipped critical height c^* , and ideally it should be set as close to \mathcal{H}_{\min} as possible. This is possible if we have information about the value \mathcal{H}_{\min} , which is the case for some statistical physics and theoretical computer science models [9, 10]. In general, however, we may not have access to the value \mathcal{H}_{\min} , and one general method is to tune the threshold parameter adaptively by setting the value at time t to be the running minimum up to time t generated by the chain. However, the resulting process becomes non-Markovian because of the adaptive tuning. Another adaptive tuning strategy is to set the value of c at time t to be $c_t = \mathcal{H}(y_t) - d$, where y_t is the proposed

state at time t generated by the proposal chain and $d \geq 0$ is a fixed number. In this way, the transition rate of the landscape-modified MH chain is always greater than or equal to that of the MH chain without landscape modification. We shall numerically investigate this tuning strategy and report positive results in Section 4.1.

As far as this paper is concerned, we assume a fixed c in all of the main theoretical results. We shall postpone to future work a systematic numerical study of investigating various choices of f and tuning strategies for c on benchmark functions, as well as a theoretical analysis of adaptively tuning c by the running minimum generated by the algorithm in an MH chain with landscape modification.

2.6. Connections between landscape modification and other acceleration techniques

In this section, we outline conceptual similarities and differences between MH with landscape modification and other common acceleration techniques in the literature for MH and simulated annealing.

2.6.1. *Catoni’s energy transformation algorithm.* Let $\alpha_1 \geq 0, \alpha_2 \geq 0, \alpha_3 > -\mathcal{H}_{\min}$ be three parameters. In Catoni [1, 2], the author introduces the energy transformation algorithm by transforming the Hamiltonian \mathcal{H} to

$$F_{\alpha_1, \alpha_2, \alpha_3}(x) := \alpha_1 \mathcal{H}(x) + \alpha_2 \log(\mathcal{H}(x) + \alpha_3).$$

Recall from Section 2.1 that MH with landscape modification can be considered as a **state-dependent** version of energy transformation if we take $f(z) = z, \alpha_3 = -c + \epsilon$, with α_1, α_2 chosen in a state-dependent manner:

$$\alpha_1(x) = \mathbf{1}_{\{\mathcal{H}(x) \leq c\}}, \quad \alpha_2(x) = \mathbf{1}_{\{\mathcal{H}(x) > c\}}.$$

Note that MH with landscape modification can give rise to other kinds of energy transformation through different choices of f ; see for example the quadratic case or the square root case in Section 2.2 or 2.3, respectively. The idea of mapping or transforming the function from \mathcal{H} to $F(\mathcal{H})$ with F being strictly increasing and concave can be dated back to R. Azencott.

2.6.2. *Preconditioning of the Hamiltonian.* Landscape modification can be understood as a **state-dependent** preconditioning of the Hamiltonian \mathcal{H} . Recall that in (2.4) we compute the acceptance–rejection probability in MH by

$$\begin{aligned} & \exp(-(\mathcal{H}_{\epsilon, c}^f(y) - \mathcal{H}_{\epsilon, c}^f(x))_+) \\ &= \begin{cases} 1 & \text{if } \mathcal{H}(y) \leq \mathcal{H}(x), \\ \exp\left(\frac{1}{\epsilon}(\mathcal{H}(y) - \mathcal{H}(x))\right) & \text{if } c \geq \mathcal{H}(y) > \mathcal{H}(x), \\ \exp\left(\frac{1}{\epsilon}(c - \mathcal{H}(x)) - \int_c^{\mathcal{H}(y)} \frac{1}{f(u-c)+\epsilon} du\right) & \text{if } \mathcal{H}(y) > c \geq \mathcal{H}(x), \\ \exp\left(-\int_{\mathcal{H}(x)}^{\mathcal{H}(y)} \frac{1}{f(u-c)+\epsilon} du\right) & \text{if } \mathcal{H}(y) > \mathcal{H}(x) > c. \end{cases} \end{aligned}$$

On the set $\{c \geq \mathcal{H}(y) > \mathcal{H}(x)\}$, the acceptance–rejection probability is the same as the original MH to allow for exploitation, while on the set $\{\mathcal{H}(y) > c \geq \mathcal{H}(x)\}$ and $\{\mathcal{H}(y) > \mathcal{H}(x) > c\}$, the acceptance–rejection probability is higher than that of the original MH to encourage

exploration of the landscape. Therefore, there is a higher transition rate of moving to other states when the algorithm is above the threshold c .

2.6.3. *Importance sampling.* In importance sampling, the target distribution is altered for possible benefits and speed-ups such as variance reduction. In landscape modification, the target distribution in MH is altered from the original Gibbs distribution $\pi^0(x) \propto e^{-\frac{1}{\epsilon}\mathcal{H}(x)}$ to $\pi_{\epsilon,c}^f(x) \propto e^{-\mathcal{H}_{\epsilon,c}^f(x)}$, while the set of stationary points is preserved in the sense that \mathcal{H} and $\mathcal{H}_{\epsilon,c}^f$ share the same set of stationary points. In importance sampling, however, the set of stationary points need not be preserved between the altered Hamiltonian and the original Hamiltonian.

2.6.4 *Quantum annealing.* In quantum annealing [3], given a target Hamiltonian \mathcal{H} and an initial Hamiltonian \mathcal{H}_{init} that is usually easy to optimize, we optimize a time-dependent function Q_t defined, for $t \in [0, T]$, by

$$Q_t(x) := A(t)\mathcal{H}_{init}(x) + B(t)\mathcal{H}(x),$$

where $A(t)$ and $B(t)$ are smooth annealing schedules that satisfy $A(T) = B(0) = 0$ and T is the total annealing time. We also choose $A(t)$ to be decreasing and $B(t)$ to be increasing on the interval $[0, T]$.

In simulated annealing with landscape modification, we also optimize a time-dependent function $\mathcal{H}_{\epsilon,c}^f$ which shares the same set of stationary points as the target \mathcal{H} . In quantum annealing, \mathcal{H}_{init} and \mathcal{H} do not necessarily share the same set of stationary points. We mention the works Del Moral and Miclo [11], Löwe [12], and Frigerio and Grillo [13] for simulated annealing with time-dependent energy function.

3. The Curie–Weiss model with landscape modification

In this section, we demonstrate the power of landscape modification by revisiting the Curie–Weiss (CW) model. With an appropriate choice of parameters, the landscape of the CW free energy is modified and the local minimum is eliminated while the global minimum is preserved on the transformed function. As a result, landscape modification convexifies the free energy from a double-well to a single-well as a function of the magnetization.

Let us first recall the setting of the CW model of a ferromagnet with external field $h \in \mathbb{R}$ and fix some notation. We shall follow the setting of Bovier and den Hollander ([14], Chapters 13–14). Let $\mathcal{X} = \{-1, 1\}^N$ be the set of possible configurations of the CW model with $N \in \mathbb{N}$. The CW Hamiltonian is given, for $\sigma = (\sigma_i)_{i=1}^N \in \mathcal{X}$, by

$$H_N(\sigma) := -\frac{1}{2N} \sum_{i,j=1}^N \sigma_i \sigma_j - h \sum_{i=1}^N \sigma_i = -\frac{N}{2} m_N(\sigma)^2 - h N m_N(\sigma) =: NE(m_N(\sigma)),$$

where $m_N(\sigma) = (1/N) \sum_{i=1}^N \sigma_i$ is the empirical magnetization. Consider the continuized Glauber dynamics by picking a node uniformly at random and flipping the sign of the selected spin, while targeting the Gibbs distribution with the CW Hamiltonian H_N at temperature ϵ . The resulting Metropolis dynamics is given by

$$P_{\epsilon,N}(\sigma, \sigma') = \begin{cases} (1/N)e^{-\frac{1}{\epsilon}(H_N(\sigma')-H_N(\sigma))_+} & \text{if } \|\sigma - \sigma'\|_1 = 2, \\ -\sum_{\eta:\eta \neq \sigma} P_{\epsilon,N}(\sigma, \eta) & \text{if } \sigma = \sigma', \\ 0 & \text{otherwise,} \end{cases}$$

where $\|\cdot\|_1$ is the l^1 norm on \mathcal{X} .

The dynamics of the empirical magnetization $(m^0(t))_{t \geq 0}$ can be described by lumping the Glauber dynamics to give

$$M_N^0(m, m') = M_{\epsilon, N}^0(m, m') = \begin{cases} \frac{1-m}{2} e^{-\frac{1}{\epsilon} N(E(m')-E(m))_+} & \text{if } m' = m + 2N^{-1}, \\ \frac{1+m}{2} e^{-\frac{1}{\epsilon} N(E(m')-E(m))_+} & \text{if } m' = m - 2N^{-1}, \\ -\sum_{m': m' \neq m} M_{\epsilon, N}^0(m, m') & \text{if } m = m', \\ 0 & \text{otherwise,} \end{cases}$$

on the state space $\Gamma_N := \{-1, -1 + 2N^{-1}, \dots, 1 - 2N^{-1}, 1\}$ with the image Gibbs distribution

$$\pi_N^0(m) = \pi_{\epsilon, N}^0(m) \propto e^{-\frac{1}{\epsilon} NE(m)} \binom{N}{\frac{1+m}{2}N} 2^{-N}, \quad m \in \Gamma_N,$$

as the stationary distribution. Note that the dependency on ϵ is suppressed in the notation of M_N^0 and π_N^0 . Define

$$I_N(m) := -\frac{1}{N} \ln \left(\binom{N}{\frac{1+m}{2}N} 2^{-N} \right),$$

$$I(m) := \frac{1}{2}(1+m) \ln(1+m) + \frac{1}{2}(1-m) \ln(1-m),$$

$$g_{\epsilon, N}(m) := E(m) + \epsilon I_N(m),$$

where $I(m)$ is the Cramér rate function for coin tossing. Then the image Gibbs distribution can be written as

$$\pi_{\epsilon, N}^0(m) \propto e^{-\frac{1}{\epsilon} N g_{\epsilon, N}(m)},$$

$$\lim_{N \rightarrow \infty} I_N(m) = I(m),$$

$$g_\epsilon(m) := \lim_{N \rightarrow \infty} g_{\epsilon, N}(m) = E(m) + \epsilon I(m).$$

The quantity g_ϵ is called the free energy of the CW model. The stationary points of g_ϵ satisfy the classical mean-field equation

$$m = \tanh\left(\frac{1}{\epsilon}(m+h)\right). \tag{3.1}$$

To seek the ground state(s) of the free energy, we consider modifying the landscape of the CW Hamiltonian from H_N to

$$E_{\epsilon, c}^f(m) := \int_d^{E(m)} \frac{1}{f((u-c)_+) + \epsilon} du, \tag{3.2}$$

$$H_{\epsilon, c, N}^f(\sigma) := N \cdot E_{\epsilon, c}^f(m(\sigma)), \tag{3.3}$$

where $d \in \mathbb{R}$ can be chosen arbitrarily, since we are only interested in the difference of $E_{\epsilon,c}^f$. The infinitesimal generator of the magnetization $(m^f(t))_{t \geq 0}$ is

$$M_{\epsilon,c,N}^f(m, m') = \begin{cases} \frac{1-m}{2} e^{-N(E_{\epsilon,c}^f(m') - E_{\epsilon,c}^f(m))_+} & \text{if } m' = m + 2N^{-1}, \\ \frac{1+m}{2} e^{-N(E_{\epsilon,c}^f(m') - E_{\epsilon,c}^f(m))_+} & \text{if } m' = m - 2N^{-1}, \\ -\sum_{m': m' \neq m} M_{\epsilon,c,N}^f(m, m') & \text{if } m = m', \\ 0 & \text{otherwise,} \end{cases}$$

with stationary distribution

$$\pi_{\epsilon,c,N}^f(m) \propto e^{-N E_{\epsilon,c,N}^f(m)} \binom{N}{\frac{1+m}{2}N} 2^{-N} = e^{-N g_{\epsilon,c,N}^f(m)}, \quad m \in \Gamma_N,$$

where

$$g_{\epsilon,c,N}^f(m) := E_{\epsilon,c}^f(m) + I_N(m).$$

By taking the limit $N \rightarrow \infty$, we find that the free energy in the landscape-modified CW model is

$$g_{\epsilon,c}^f(m) := E_{\epsilon,c}^f(m) + I(m).$$

Setting the derivative of $g_{\epsilon,c}^f$ equal to zero gives the **landscape-modified mean-field equation**:

$$m = \tanh\left(\frac{m+h}{f((E(m)-c)_+) + \epsilon}\right). \tag{3.4}$$

Observe that if we take $f = 0$, then (3.4) reduces to the classical mean-field equation in (3.1).

3.1. Main results

Without loss of generality, assume the external magnetic field is $h < 0$. In the subcritical regime where $\frac{1}{\epsilon} > 1$, it is known that there are two local minima of g_ϵ . We denote the global minimum of g_ϵ by $m_-^* < 0$ and the other local minimum by $m_+^* > 0$, where $|m_-^*| > m_+^*$, and let z^* be the saddle point between m_-^* and m_+^* . We also write $m_-^*(N)$ (resp. $m_+^*(N)$) for the closest point in Euclidean distance to m_-^* (resp. m_+^*) on Γ_N .

Theorem 1. (Landscape modification in the subcritical regime.) *Suppose $\frac{1}{\epsilon} > 1$, $h < 0$, and f, c are chosen as in Assumption 1.*

1. (Convexification of the free energy g_ϵ and subexponential mean crossover time.) *If we choose $c \in [E(m_-^*), E(m_+^*)]$, $c < h^2/2$, $-h - \sqrt{h^2 - 2c} \leq z^*$, and for $m \in [-h - \sqrt{h^2 - 2c}, -h + \sqrt{h^2 - 2c}]$,*

$$m > \tanh\left(\frac{m+h}{f((E(m)-c)_+) + \epsilon}\right),$$

then m_-^* is the only stationary point of the modified free energy $g_{\epsilon,c}^f$, which is a global minimum. Consequently, we have subexponential mean crossover time on the modified landscape

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{E}_{m_+^*(N)} \left(\tau_{m_-^*(N)}^f \right) = 0, \tag{3.5}$$

while the mean crossover time on the original landscape is exponential in N , $1/\epsilon$, and the original critical height $g_\epsilon(z^*) - g_\epsilon(m_+^*)$ with

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{E}_{m_+^*(N)} \left(\tau_{m_-^*(N)}^0 \right) = \frac{1}{\epsilon} (g_\epsilon(z^*) - g_\epsilon(m_+^*)).$$

2. If we choose $c \in [E(m_+^*), E(z^*)]$ and assume in addition that f is twice differentiable and satisfies $f'(0) = f''(0) = 0$, then there exists

$$z^* = \arg \max_{m_-^* \leq m \leq m_+^*} g_{\epsilon,c}^f(m),$$

and as $N \rightarrow \infty$,

$$\begin{aligned} \mathbb{E}_{m_+^*(N)} \left(\tau_{m_-^*(N)}^f \right) &= \exp \left(N \left(g_{\epsilon,c}^f(z^*) - g_{\epsilon,c}^f(m_+^*) \right) \right) \\ &\times \frac{2}{1 - |z^*|} \sqrt{\frac{1 - z^{*2}}{1 - m_+^{*2}}} \frac{2\pi N/4}{\sqrt{\left(- \left(g_{\epsilon,c}^f \right)''(z^*) \right) \left(g_{\epsilon,c}^f \right)''(m_+^*)}} (1 + o(1)). \end{aligned}$$

Consequently,

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{E}_{m_+^*(N)} \left(\tau_{m_-^*(N)}^f \right) &= g_{\epsilon,c}^f(z^*) - g_{\epsilon,c}^f(m_+^*) \\ &= \frac{1}{\epsilon} (c - E(m_+^*)) + \int_c^{E(z^*)} \frac{1}{f((u - c)_+) + \epsilon} du \\ &\quad + (I(z^*) - I(m_+^*)) \\ &\leq \frac{1}{\epsilon} (g_\epsilon(z^*) - g_\epsilon(m_+^*)) = \lim_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{E}_{m_+^*(N)} \left(\tau_{m_-^*(N)}^0 \right). \end{aligned}$$

Before we present the proof, we interpret the results in Theorem 1 intuitively: in Item 1, on the one hand we would like to choose c small enough so that the mapping

$$m \mapsto \tanh \left(\frac{m + h}{f((E(m) - c)_+) + \epsilon} \right)$$

is flattened and only intersects the straight line $m \mapsto m$ at the global minimum m_-^* . In this way the landscape of $g_{\epsilon,c}^f$ is transformed from a double-well to a single-well landscape, while the location of the global minimum at m_-^* is preserved as that in the original landscape g_ϵ . This is illustrated in Figures 2 and 3. On the other hand, we cannot choose c to be too small if we are interested in seeking the ground state of g_ϵ , since otherwise, if $c < E(m_-^*)$, then m_-^* may no

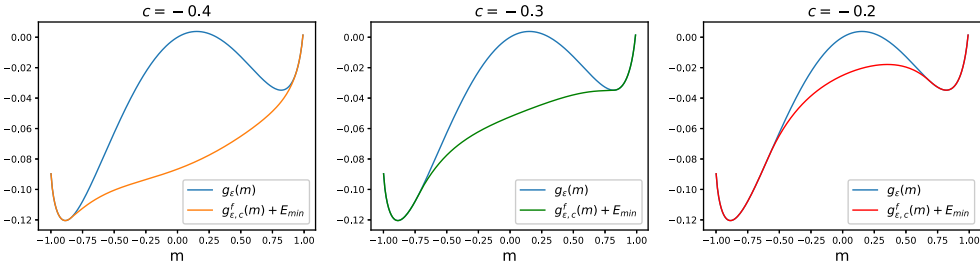


FIGURE 2. Plots of the free energy g_ϵ and the modified free energy $g_{\epsilon,c}^f + E_{min}$ with $h = -0.05$ and $f(z) = z$ at temperature $\epsilon = 1/1.5$, where $E_{min} = \min_{m \in [-1, 1]} E(m)$. We shift the modified free energy by E_{min} so that it is on the same scale as the original free energy g_ϵ .

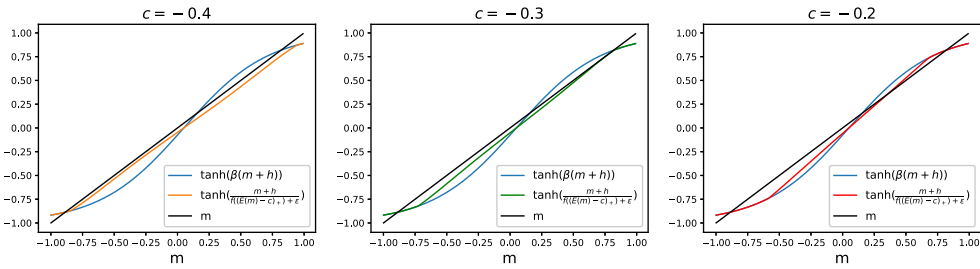


FIGURE 3. Plots of the mean-field equation (3.1) and the modified mean-field equation (3.4) with $h = -0.05$ and $f(z) = z$ at temperature $\epsilon = 1/1.5$.

longer be the global minimum in the transformed free energy $g_{\epsilon,c}^f$. This consequently yields a mean crossover time on the modified landscape that is subexponential in N , while the original mean crossover time is exponential in N , $1/\epsilon$, and the original critical height $g_\epsilon(z^*) - g_\epsilon(m_+^*)$. In Item 2 of Theorem 1, we choose a larger value of c compared with that in Item 1. Although the transformed free energy $g_{\epsilon,c}^f$ is not a convex function, it has a smaller critical height than the original free energy g_ϵ . This gives a reduced exponential dependence on the modified mean crossover time compared with the original mean crossover time.

The power of landscape modification or energy transformation lies in tuning the parameter c appropriately. One way to tune c is to use the running minimum generated by the algorithm on the original free energy g_ϵ . Suppose we start in the well containing the local minimum m_+^* ; setting c to be the running minimum eventually gives $c = E(m_+^*)$, and hence Item 2 of Theorem 1 can be applied and the critical height on the modified landscape is reduced.

We illustrate Theorem 1 with a concrete numerical example in Figures 2 and 3, where we take $h = -0.05$ and $f(z) = z$ at temperature $\epsilon = 1/1.5$. We numerically compute that $m_-^* = -0.8863$, $m_+^* = 0.8188$, and $z^* = 0.1524$. As a result we have $E(m_-^*) = -0.4371$, $E(m_+^*) = -0.2943$, and $E(z^*) = -0.004$. In the leftmost plots of Figures 2 and 3, we choose $c = -0.4 \in [E(m_-^*), E(m_+^*)]$. We numerically check that the conditions in Item 1 of Theorem 1 are satisfied, and we see that the location of the global minimum is the same for the blue curve as for the orange curve. In the rightmost plots of Figures 2 and 3, we choose $c = -0.2 \in [E(m_+^*), E(z^*)]$. We see that for the blue curve and the red curve, the locations of the two local minima are the same, while the critical height is smaller than that in the original landscape g_ϵ .

3.2. Proof of Theorem 1

Before we give the proof, let us first recall the concept of critical height and the notation L as introduced in Section 2. In this section, we are interested in the CW model with and without landscape modification, with free energy $g_{\epsilon,c,N}^f$ and $g_{\epsilon,N}$ respectively. We therefore define the analogous concepts of critical height by inserting a subscript of N . This leads us to

$$H_{\epsilon,c,N}^f := L(g_{\epsilon,c,N}^f) = \max_{x,y \in \mathcal{X}} \left\{ G(g_{\epsilon,c,N}^f, x, y) - g_{\epsilon,c,N}^f(x) - g_{\epsilon,c,N}^f(y) \right\} + \min_{x \in \Gamma_N} g_{\epsilon,c,N}^f(x),$$

$$H_{\epsilon,N}^0 := L(g_{\epsilon,N}) = \max_{x,y \in \mathcal{X}} \left\{ G(g_{\epsilon,N}, x, y) - g_{\epsilon,N}(x) - g_{\epsilon,N}(y) \right\} + \min_{x \in \Gamma_N} g_{\epsilon,N}(x),$$

where we recall that G is introduced in (2.3).

We now proceed with the proof.

Proof of Theorem 1. First we prove Item 1. We observe that $\{E(m) \geq c\} = \{m \in [-h - \sqrt{h^2 - 2c}, -h + \sqrt{h^2 - 2c}]\}$. On this interval,

$$\frac{d}{dm} g_{\epsilon,c}^f(m) = \operatorname{arctanh}(m) - \frac{m+h}{f((E(m)-c)_+) + \epsilon} > 0,$$

and hence the modified free energy is strictly increasing on this interval. On the interval $\{m > -h + \sqrt{h^2 - 2c}\}$,

$$\frac{d}{dm} g_{\epsilon,c}^f(m) = \frac{d}{dm} g_{\epsilon}(m) > 0,$$

as the original free energy is strictly increasing. On the interval $\{m < -h - \sqrt{h^2 - 2c}\}$, we also have

$$\frac{d}{dm} g_{\epsilon,c}^f(m) = \frac{d}{dm} g_{\epsilon}(m).$$

Thus, with these parameter choices, the only stationary point of $g_{\epsilon,c}^f$ is m^* , which is the global minimum.

Next we proceed to proving (3.5). According to Löwe ([12], Theorem 2.1), for $\xi_1(N)$ a polynomial function in N , we have

$$\frac{1}{\lambda_2(-M_{\epsilon,c,N}^f)} \leq \xi_1(N) e^{NH_{\epsilon,c,N}^f}.$$

Using the random target lemma (Aldous and Fill, [15], Section 4.2) and the above inequality leads to

$$\begin{aligned} \pi_{\epsilon,c,N}^f(m_{-}^*(N)) \mathbb{E}_{m_{+}^*(N)} \left(\tau_{m_{-}^*(N)}^f \right) &\leq \sum_{y \in \Gamma_N} \pi_{\epsilon,c,N}^f(y) \mathbb{E}_{m_{+}^*(N)} \left(\tau_y^f \right) \leq (|\Gamma_N| - 1) \frac{1}{\lambda_2(-M_{\epsilon,c,N}^f)} \\ &\leq N \xi_1(N) e^{NH_{\epsilon,c,N}^f}. \end{aligned} \tag{3.6}$$

Now, let $\bar{m}(N) := \arg \min g_{\epsilon,c,N}^f(m)$ and compute

$$\begin{aligned} g_{\epsilon,c,N}^f(m_-^*(N)) - g_{\epsilon,c,N}^f(\bar{m}(N)) &= E_{\epsilon,c}^f(m_-^*(N)) - E_{\epsilon,c}^f(\bar{m}(N)) + I_N(m_-^*(N)) - I_N(\bar{m}(N)) \\ &= E_{\epsilon,c}^f(m_-^*(N)) - E_{\epsilon,c}^f(\bar{m}(N)) \\ &\quad + I_N(m_-^*(N)) - I(m_-^*(N)) + I(m_-^*(N)) - I_N(\bar{m}(N)) \\ &= E_{\epsilon,c}^f(m_-^*(N)) - E_{\epsilon,c}^f(\bar{m}(N)) \\ &\quad + [1 + o(1)] \frac{1}{2N} \ln \left(\frac{\pi N (1 - m_-^*(N)^2)}{2} \right) \\ &\quad + I(m_-^*(N)) - I(\bar{m}(N)) \\ &\quad - [1 + o(1)] \frac{1}{2N} \ln \left(\frac{\pi N (1 - \bar{m}(N)^2)}{2} \right) \\ &\rightarrow 0 \quad \text{as } N \rightarrow \infty, \end{aligned}$$

where we use Bovier and den Hollander ([14], Equation (13.2.5)) in the third equality, and $m_-^*(N), \bar{m}(N) \rightarrow m_-^*$ as $N \rightarrow \infty$. The above computation combined with (3.6) yields

$$\limsup_{N \rightarrow \infty} \mathbb{E}_{m_+^*(N)} \left(\tau_{m_-^*(N)}^f \right) \leq \lim_{N \rightarrow \infty} \frac{1}{N} \log H_{\epsilon,c,N}^f = 0. \tag{3.7}$$

On the other hand, as the magnetization $(m^f(t))_{t \geq 0}$ is a birth–death process, using Equation (13.2.2) of Bovier and den Hollander [14], the mean hitting time can be calculated explicitly as

$$\begin{aligned} &\mathbb{E}_{m_+^*(N)} \left(\tau_{m_-^*(N)}^f \right) \\ &= \sum_{\substack{m,m' \in \Gamma_N, m \leq m' \\ m_-^*(N) < m \leq m_+^*(N)}} \frac{\pi_{\epsilon,c,N}^f(m')}{\pi_{\epsilon,c,N}^f(m)} \frac{1}{M_{\epsilon,c,N}^f(m, m - 2N^{-1})} \\ &\geq \frac{1}{M_{\epsilon,c,N}^f(m_+^*(N), m_+^*(N) - 2N^{-1})} = \frac{2}{1 + m_+^*(N)} e^{N(E_{\epsilon,c}^f(m_+^*(N)) - 2N^{-1}) - E_{\epsilon,c}^f(m)_+} \\ &\geq \frac{2}{1 + m_+^*(N)}. \end{aligned}$$

As a result, as $m_+^*(N) \rightarrow m_+^*$ we have

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{E}_{m_+^*(N)} \left(\tau_{m_-^*(N)}^f \right) \geq 0. \tag{3.8}$$

Using both (3.8) and (3.7) gives (3.5).

Next, we prove Item 2, the argument for which closely follows that of Bovier and den Hollander ([14], Theorem 13.1). Since we choose $c \in [E(m_+^*), E(z^*)]$, we have $E(m_-^*) - c <$

$E(m_+^*) - c \leq 0$, and hence the landscape-modified mean-field equation (3.4) has at least two solutions m_+^* and m_-^* , which are exactly the same as those of the original mean-field equation (3.1). As the landscape-modified mean-field equation is continuous in m , there exists $\mathbf{z}^* = \arg \max_{m_-^* \leq m \leq m_+^*} g_{\epsilon,c}^f(m)$ which also satisfies (3.4). Now, for $m \in \Gamma_N$ we consider

$$N(E_{\epsilon,c}^f(m - 2N^{-1}) - E_{\epsilon,c}^f(m)) = N \int_{E(m)}^{E(m)+2N^{-1}(m+h-N^{-1})} \frac{1}{f((u-c)_+) + \epsilon} du$$

$$\rightarrow \frac{2(m+h)}{f((E(m)-c)_+) + \epsilon} \text{ as } N \rightarrow \infty.$$

If we take $N \rightarrow \infty$ and $m \rightarrow \mathbf{z}^*$, we obtain

$$\frac{1}{M_{\epsilon,c,N}^f(m, m - 2N^{-1})} \rightarrow \frac{2}{1 + \mathbf{z}^*} \exp\left(\frac{2(\mathbf{z}^* + h)_+}{f((E(\mathbf{z}^*) - c)_+) + \epsilon}\right) = \frac{2}{1 - |\mathbf{z}^*|},$$

since, if $\mathbf{z}^* > 0$, then $\mathbf{z}^* + h > 0$ and satisfies the mean-field equation (3.4). Using the mean hitting time formula again leads to the following, for any $\delta > 0$:

$$\mathbb{E}_{m_+^*(N)}(\tau_{m_-^*(N)}^f) = \sum_{\substack{m,m' \in \Gamma_N, m \leq m' \\ m_-^*(N) < m \leq m_+^*(N)}} \frac{\pi_{\epsilon,c,N}^f(m')}{\pi_{\epsilon,c,N}^f(m)} \frac{1}{M_{\epsilon,c,N}^f(m, m - 2N^{-1})}$$

$$= e^{N[g_{\epsilon,c,N}^f(\mathbf{z}^*) - g_{\epsilon,c,N}^f(m_+^*)]} \frac{2}{1 - |\mathbf{z}^*|} [1 + o(1)]$$

$$\times \sum_{\substack{m,m' \in \Gamma_N \\ |m - \mathbf{z}^*| < \delta, |m' - m_+^*| < \delta}} e^{N[g_{\epsilon,c,N}^f(m) - g_{\epsilon,c,N}^f(\mathbf{z}^*)] - N[g_{\epsilon,c,N}^f(m') - g_{\epsilon,c,N}^f(m_+^*)]}. \quad (3.9)$$

Using Stirling’s formula and the same argument as in Bovier and den Hollander ([14], Equation (13.2.7)) yields

$$e^{N[g_{\epsilon,c,N}^f(\mathbf{z}^*) - g_{\epsilon,c,N}^f(m_+^*)]} = (1 + o(1)) e^{N[g_{\epsilon,c}^f(\mathbf{z}^*) - g_{\epsilon,c}^f(m_+^*)]} \sqrt{\frac{1 - \mathbf{z}^{*2}}{1 - m_+^{*2}}}.$$

Substitution into (3.9) gives

$$\mathbb{E}_{m_+^*(N)}(\tau_{m_-^*(N)}^f)$$

$$= e^{N[g_{\epsilon,c}^f(\mathbf{z}^*) - g_{\epsilon,c}^f(m_+^*)]} \frac{2}{1 - |\mathbf{z}^*|} \sqrt{\frac{1 - \mathbf{z}^{*2}}{1 - m_+^{*2}}} [1 + o(1)]$$

$$\times \sum_{\substack{m,m' \in \Gamma_N \\ |m - \mathbf{z}^*| < \delta, |m' - m_+^*| < \delta}} \sqrt{\frac{1 - m^2}{1 - \mathbf{z}^{*2}}} \sqrt{\frac{1 - m_+^{*2}}{1 - m'^2}} e^{N[g_{\epsilon,c}^f(m) - g_{\epsilon,c}^f(\mathbf{z}^*)] - N[g_{\epsilon,c}^f(m') - g_{\epsilon,c}^f(m_+^*)]}. \quad (3.10)$$

We use a Laplace method argument to handle the sum in (3.10). Note that as f is assumed to be twice differentiable with $f(0) = f'(0) = f''(0) = 0$, this implies that $g_{\epsilon,c}^f$ is three times differentiable, and applying the third-order Taylor expansion gives

$$g_{\epsilon,c}^f(m) - g_{\epsilon,c}^f(\mathbf{z}^*) = \frac{(m - \mathbf{z}^*)^2}{2} (g_{\epsilon,c}^f)''(\mathbf{z}^*) + \mathcal{O}\left((m - \mathbf{z}^*)^3\right), \tag{3.11}$$

$$g_{\epsilon,c}^f(m') - g_{\epsilon,c}^f(m_+^*) = \frac{(m' - m_+^*)^2}{2} (g_{\epsilon,c}^f)''(m_+^*) + \mathcal{O}\left((m' - m_+^*)^3\right), \tag{3.12}$$

where we use $(g_{\epsilon,c}^f)'(\mathbf{z}^*) = (g_{\epsilon,c}^f)'(m_+^*) = 0$. Now we observe that the sum in (3.10) is

$$\begin{aligned} & (1 + o(1)) \frac{N}{4} \int_{\mathbb{R}} \int_{\mathbb{R}} \exp\left[\frac{1}{2} (g_{\epsilon,c}^f)''(\mathbf{z}^*) u^2 - \frac{1}{2} (g_{\epsilon,c}^f)''(m_+^*) u'^2\right] dud u' \\ &= (1 + o(1)) \frac{N}{4} \frac{2\pi}{\sqrt{\left[-(g_{\epsilon,c}^f)''(\mathbf{z}^*)\right] (g_{\epsilon,c}^f)''(m_+^*)}}, \end{aligned}$$

since $(g_{\epsilon,c}^f)''(\mathbf{z}^*) < 0$ and $(g_{\epsilon,c}^f)''(m_+^*) > 0$. □

3.3. Extension to the random-field Curie–Weiss model

In Section 3.1, we discussed the classical CW model with landscape modification under a fixed magnetic field. In this section, we consider the random-field CW model with landscape modification. We discuss related metastability results and the ground-state free energy in such a setting, with the aim of illustrating that landscape modification can also be applied in the setting of a random energy landscape. Let us begin by recalling the random-field CW model. We shall adapt the setting of Mathieu and Picco [16]. Let $(h_i)_{i \in \mathbb{N}}$ be a sequence of independent and identically distributed random variables with $\mathbb{P}(h_i = 1) = \mathbb{P}(h_i = -1) = 1/2$. We consider the random Hamiltonian function given, for a fixed $\theta > 0$ and $\sigma \in \{-1, 1\}^N$, by

$$\begin{aligned} \mathbf{H}_N(\sigma) = \mathbf{H}_N(\sigma, \omega) &:= -\frac{N}{2} \mathbf{m}_N(\sigma)^2 - \theta \sum_{i=1}^N h_i(\omega) \sigma_i \\ &= -\frac{N}{2} (\mathbf{m}_N^+(\sigma) + \mathbf{m}_N^-(\sigma))^2 - \theta N (\mathbf{m}_N^+(\sigma) - \mathbf{m}_N^-(\sigma)), \end{aligned}$$

where

$$\mathbf{m}_N(\sigma) := (1/N) \sum_{i=1}^N \sigma_i, \quad \mathbf{m}_N^+(\sigma) := (1/N) \sum_{i=1; h_i=1}^N \sigma_i, \quad \mathbf{m}_N^-(\sigma) := (1/N) \sum_{i=1; h_i=-1}^N \sigma_i.$$

In the sequel, we shall suppress the dependency on ω . Denote the Gibbs distribution at temperature $\epsilon \in \{-1, 1\}^N$ by

$$\mathbf{v}_N(\sigma) \propto \exp\left\{-\frac{1}{\epsilon} \mathbf{H}_N(\sigma)\right\}.$$

Let $N^+ := |\{i; h_i = +1\}|$, $N^- := |\{i; h_i = -1\}|$, and define the random set

$$\Gamma_N := \left(-\frac{N^+}{N}, -\frac{N^+}{N} + \frac{2}{N}, \dots, \frac{N^+}{N}\right) \times \left(-\frac{N^-}{N}, -\frac{N^-}{N} + \frac{2}{N}, \dots, \frac{N^-}{N}\right).$$

For $\mathbf{m} = (\mathbf{m}^+, \mathbf{m}^-) \in \Gamma_N$, with slight abuse of notation we write

$$\mathbf{H}_N(\mathbf{m}) = -\frac{N}{2}(\mathbf{m}^+ + \mathbf{m}^-)^2 - \theta N(\mathbf{m}^+ - \mathbf{m}^-) =: N \cdot \mathbf{E}(\mathbf{m}),$$

where \mathbf{m}^+ and \mathbf{m}^- are the magnetization among the sites i where respectively $h_i = 1$ and $h_i = -1$. Let $\pi_{\epsilon, N}^0$ denote the image Gibbs distribution of ν_N by Γ_N , where

$$\pi_{\epsilon, N}^0(\mathbf{m}) \propto \exp \left\{ -\frac{1}{\epsilon} N \mathbf{g}_{\epsilon, N}(\mathbf{m}) \right\},$$

$$\mathbf{g}_{\epsilon, N}(\mathbf{m}) := -\frac{1}{2}(\mathbf{m}^+ + \mathbf{m}^-)^2 - \theta(\mathbf{m}^+ - \mathbf{m}^-) - \frac{1}{\frac{1}{\epsilon}N} \log \left(\frac{N^+}{\frac{N^+}{2} + \mathbf{m}^+ \frac{N}{2}} \right) \left(\frac{N^-}{\frac{N^-}{2} + \mathbf{m}^- \frac{N}{2}} \right).$$

As $N \rightarrow \infty$, by the strong law of large numbers, $\mathbf{g}_{\epsilon, N}$ converges almost surely to the free energy given by

$$\mathbf{g}_\epsilon(\mathbf{m}) := -\frac{1}{2}(\mathbf{m}^+ + \mathbf{m}^-)^2 - \theta(\mathbf{m}^+ - \mathbf{m}^-) + \frac{1}{2\frac{1}{\epsilon}} (I(2\mathbf{m}^+) + I(2\mathbf{m}^-)),$$

where $I(m)$ is the Cramér rate function as introduced in Section 3. The critical points of \mathbf{g}_ϵ satisfy

$$\mathbf{m}^+ = \frac{1}{2} \tanh \left(\frac{1}{\epsilon} (\mathbf{m}^+ + \mathbf{m}^- + \theta) \right), \tag{3.13}$$

$$\mathbf{m}^- = \frac{1}{2} \tanh \left(\frac{1}{\epsilon} (\mathbf{m}^+ + \mathbf{m}^- - \theta) \right). \tag{3.14}$$

In this section, we shall only consider the subcritical regime where $\frac{1}{\epsilon} > \cosh^2(\frac{1}{\epsilon}\theta)$. It can be shown (see e.g. [16]) that there are exactly three critical points. Let $\mathbf{m}_* > 0$ be the unique positive solution to the mean-field equation

$$\mathbf{m}_* = \frac{1}{2} \left(\tanh \left(\frac{1}{\epsilon} (\mathbf{m}_* + \theta) \right) + \tanh \left(\frac{1}{\epsilon} (\mathbf{m}_* - \theta) \right) \right).$$

The three critical points of \mathbf{g}_ϵ are given by

$$\mathbf{m}_0 = \left(\frac{1}{2} \tanh \left(\frac{1}{\epsilon} \theta \right), -\frac{1}{2} \tanh \left(\frac{1}{\epsilon} \theta \right) \right),$$

$$\mathbf{m}_1 = \left(\frac{1}{2} \tanh \left(\frac{1}{\epsilon} \mathbf{m}_* + \frac{1}{\epsilon} \theta \right), \frac{1}{2} \tanh \left(\frac{1}{\epsilon} \mathbf{m}_* - \frac{1}{\epsilon} \theta \right) \right),$$

$$\mathbf{m}_2 = \left(\frac{1}{2} \tanh \left(-\frac{1}{\epsilon} \mathbf{m}_* + \frac{1}{\epsilon} \theta \right), -\frac{1}{2} \tanh \left(\frac{1}{\epsilon} \mathbf{m}_* + \frac{1}{\epsilon} \theta \right) \right),$$

where \mathbf{m}_0 is the saddle point and $\mathbf{m}_1, \mathbf{m}_2$ are the two global minima. Consider the continuized Glauber dynamics $(\sigma_N(t))_{t \geq 0}$ by picking a node uniformly at random and changing the sign of the selected spin, while targeting the Gibbs distribution ν_N at temperature ϵ . Denote by $\mathbf{m}_N(t) := \mathbf{m}_N(\sigma_N(t))$ the induced dynamics on the magnetization, and denote its infinitesimal generator by $\mathbf{M}_{\epsilon,N}^0$. This is proven to be a Markov chain in Mathieu and Picco [16], with stationary measure $\pi_{\epsilon,N}^0$.

Now let us consider the landscape-modified Hamiltonian on Γ_N ,

$$\mathbf{E}_{\epsilon,c}^f(\mathbf{m}) := \int_d^{\mathbf{E}(\mathbf{m})} \frac{1}{f((u-c)_+) + \epsilon} du, \tag{3.15}$$

$$\mathbf{H}_{\epsilon,c,N}^f(\mathbf{m}) := N \cdot \mathbf{E}_{\epsilon,c}^f(\mathbf{m}), \tag{3.16}$$

where $d \in \mathbb{R}$ can be chosen arbitrarily since we are only interested in the difference of $\mathbf{E}_{\epsilon,c}^f$. The transformed image Gibbs distribution is therefore

$$\begin{aligned} \pi_{\epsilon,c,N}^f(\mathbf{m}) &\propto \exp\{-N\mathbf{g}_{\epsilon,c,N}^f(\mathbf{m})\}, \\ \mathbf{g}_{\epsilon,c,N}^f(\mathbf{m}) &:= \mathbf{E}_{\epsilon,c}^f(\mathbf{m}) - \frac{1}{N} \log \left(\binom{N^+}{\frac{N^+}{2} + \mathbf{m}^+ \frac{N}{2}} \binom{N^-}{\frac{N^-}{2} + \mathbf{m}^- \frac{N}{2}} \right). \end{aligned}$$

The strong law of large numbers yields that as $N \rightarrow \infty$, $\mathbf{g}_{\epsilon,c,N}^f$ converges almost surely to the transformed free energy

$$\mathbf{g}_{\epsilon,c}^f(\mathbf{m}) := \mathbf{E}_{\epsilon,c}^f(\mathbf{m}) + \frac{1}{2} (I(2\mathbf{m}^+) + I(2\mathbf{m}^-)).$$

The critical points of $\mathbf{g}_{\epsilon,c}^f$ satisfy the following **landscape-modified mean-field equations**:

$$\mathbf{m}^+ = \frac{1}{2} \tanh\left(\frac{\mathbf{m}^+ + \mathbf{m}^- + \theta}{f((\mathbf{E}(\mathbf{m}) - c)_+) + \epsilon}\right), \tag{3.17}$$

$$\mathbf{m}^- = \frac{1}{2} \tanh\left(\frac{\mathbf{m}^+ + \mathbf{m}^- - \theta}{f((\mathbf{E}(\mathbf{m}) - c)_+) + \epsilon}\right). \tag{3.18}$$

Note that (3.17) and (3.18) reduce to the classical case (3.13) and (3.14) if we take $f = 0$. Consider the continuized Glauber dynamics $(\sigma_N^f(t))_{t \geq 0}$ by picking a node uniformly at random and changing the sign of the selected spin, while targeting the Gibbs distribution with Hamiltonian $\epsilon \mathbf{H}_{\epsilon,c,N}^f$ at temperature ϵ . Denote by $\mathbf{m}_N^f(t) := \mathbf{m}_N^f(\sigma_N^f(t))$ the induced dynamics on the magnetization, and denote by $\mathbf{M}_{\epsilon,c,N}^f$ its infinitesimal generator, which is a Markov chain with stationary measure $\pi_{\epsilon,c,N}^f$.

In the following, we shall consider the case where $c \in [\mathbf{E}(\mathbf{m}_1), \mathbf{E}(\mathbf{m}_0)]$. It can be seen that the two global minima of $\mathbf{g}_{\epsilon,c}^f$ remain at $\mathbf{m}_1, \mathbf{m}_2$ with this choice of c . For any path $\gamma^{\mathbf{m}_1, \mathbf{m}_0}$ connecting \mathbf{m}_1 and \mathbf{m}_0 , we define

$$\begin{aligned} \mathbf{m}_3(\gamma^{\mathbf{m}_1, \mathbf{m}_0}) &:= \arg \max\{\mathbf{g}_{\epsilon,c}^f(\gamma_i); \gamma_i \in \gamma^{\mathbf{m}_1, \mathbf{m}_0}\}, \\ \Delta \mathbf{g}_{\epsilon,c}^f &:= \min_{\gamma^{\mathbf{m}_1, \mathbf{m}_0}} \mathbf{m}_3(\gamma^{\mathbf{m}_1, \mathbf{m}_0}) - \mathbf{g}_{\epsilon,c}^f(\mathbf{m}_1) = \min_{\gamma^{\mathbf{m}_1, \mathbf{m}_0}} \mathbf{m}_3(\gamma^{\mathbf{m}_1, \mathbf{m}_0}) - \mathbf{g}_{\epsilon,c}^f(\mathbf{m}_0), \end{aligned}$$

where $\Delta \mathbf{g}_{\epsilon,c}^f$ is the critical height on the modified landscape. We also write $\Delta \mathbf{g}_{\epsilon}$ to denote the critical height on the original landscape. Suppose that $\Delta \mathbf{g}_{\epsilon}$ is attained at \mathbf{m}_4 , so that $\Delta \mathbf{g}_{\epsilon} = \mathbf{g}_{\epsilon}(\mathbf{m}_4) - \mathbf{g}_{\epsilon}(\mathbf{m}_0)$; we deduce

$$\begin{aligned} \Delta \mathbf{g}_{\epsilon,c}^f &\leq \mathbf{g}_{\epsilon,c}^f(\mathbf{m}_4) - \mathbf{g}_{\epsilon,c}^f(\mathbf{m}_0) = \int_{\mathbf{E}(\mathbf{m}_0)}^{\mathbf{E}(\mathbf{m}_4)} \frac{1}{f((u-c)_+) + \epsilon} du \\ &\quad + \frac{1}{2}(I(2\mathbf{m}_4^+) + I(2\mathbf{m}_4^-)) - \frac{1}{2}(I(2\mathbf{m}_0^+) + I(2\mathbf{m}_0^-)) \\ &\leq \frac{1}{\epsilon} (\mathbf{g}_{\epsilon}(\mathbf{m}_4) - \mathbf{g}_{\epsilon}(\mathbf{m}_0)) = \frac{1}{\epsilon} \Delta \mathbf{g}_{\epsilon}. \end{aligned}$$

In other words, the critical height of the free energy in the modified landscape is bounded above by $\frac{1}{\epsilon}$ times the critical height of the free energy in the original landscape.

A direct application of Mathieu and Picco ([16], Theorem 2.7) yields the following result on the asymptotics of the spectral gap.

Theorem 2. (Asymptotics of the spectral gap.) *Suppose $\theta > 0$, $\frac{1}{\epsilon} > \cosh^2\left(\frac{1}{\epsilon}\theta\right)$ are fixed, and \mathbf{m}_0 is the saddle point, while $\mathbf{m}_1, \mathbf{m}_2$ are the two global minima on the original free energy landscape \mathbf{g}_{ϵ} . For $c \in [\mathbf{E}(\mathbf{m}_1), \mathbf{E}(\mathbf{m}_0)]$, we have \mathbb{P} -almost surely that*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log \lambda_2(-\mathbf{M}_{\epsilon,c,N}^f) = -\Delta \mathbf{g}_{\epsilon,c}^f \geq -\frac{1}{\epsilon} \Delta \mathbf{g}_{\epsilon} = \lim_{N \rightarrow \infty} \frac{1}{N} \log \lambda_2(-\mathbf{M}_{\epsilon,N}^0).$$

In essence, the relaxation time in the mean-field limit of the transformed generator $\mathbf{M}_{\epsilon,c,N}^f$ is asymptotically less than or equal to that of the original generator $\mathbf{M}_{\epsilon,N}^0$.

This yields a reduced exponential dependence of the relaxation time on the modified landscape, compared with the relaxation time on the original landscape.

4. Discrete simulated annealing with landscape modification

Unlike in previous sections of this paper, where the temperature parameter is fixed, in this section we consider the non-homogeneous MH with landscape modification where the temperature schedule $(\epsilon_t)_{t \geq 0}$ is time-dependent and non-increasing and goes to zero as $t \rightarrow \infty$.

We first recall the concept of critical height as introduced in Section 2. To be precise, we define

$$H_{\epsilon,c}^f := L(\mathcal{H}_{\epsilon,c}^f) = \max_{x,y \in \mathcal{X}} \left\{ G(\mathcal{H}_{\epsilon,c}^f, x, y) - \mathcal{H}_{\epsilon,c}^f(x) - \mathcal{H}_{\epsilon,c}^f(y) \right\}, \tag{4.1}$$

$$H^0 := L(\mathcal{H}) = \max_{x,y \in \mathcal{X}} \{ G(\mathcal{H}, x, y) - \mathcal{H}(x) - \mathcal{H}(y) \} + \min \mathcal{H}, \tag{4.2}$$

$$c^* := c^*(\mathcal{H}, c) = \max_{x,y \in \mathcal{X}} \{ (G(\mathcal{H}, x, y) \wedge c) - (\mathcal{H}(x) \wedge c) - (\mathcal{H}(y) \wedge c) \} + \min \mathcal{H}, \tag{4.3}$$

where $H_{\epsilon,c}^f$ is the critical height associated with the modified landscape, H^0 is the critical height associated with the original landscape \mathcal{H} , and c^* is the clipped critical height. We shall see in the main results of this section that both $H_{\epsilon,c}^f$ and c^* play fundamental roles in the relaxation

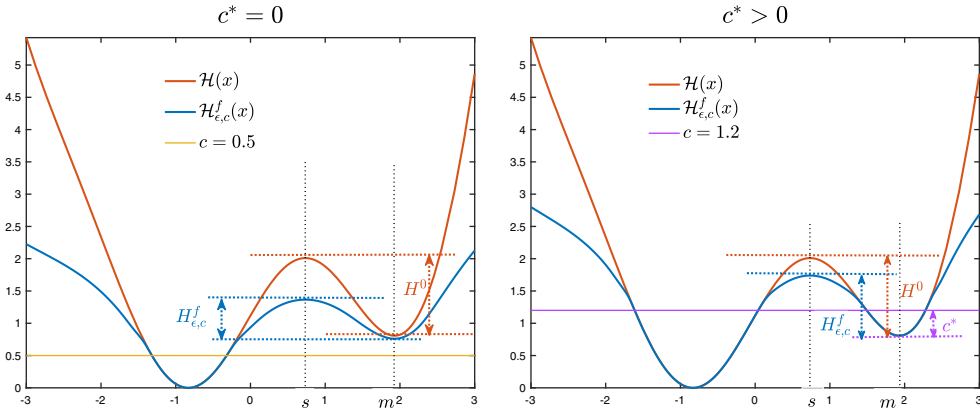


FIGURE 4. Comparing the critical heights $H^0, H_{1,c}^f, c^*$ on a one-dimensional landscape generated by \mathcal{H} with a local minimum at m and a global minimum, where we take $\epsilon = 1, f(z) = z$, and the proposal chain is of nearest-neighbour type, i.e. going left or right with probability $1/2$.

time in the low-temperature regime and in determining the cooling schedule of an improved simulated annealing algorithm running on the modified landscape.

As an illustration to calculate and compare these critical heights, we consider a simple one-dimensional landscape with a saddle point at s , a local (but not global) minimum at m , and a single global minimum. At temperature $\epsilon = 1$, the original critical height is attained at $H^0 = \mathcal{H}(s) - \mathcal{H}(m)$ while the modified critical height is $H_{1,c}^f = \mathcal{H}_{1,c}^f(s) - \mathcal{H}_{1,c}^f(m) \leq H^0$. In this setting, depending on whether c is above or below $\mathcal{H}(m)$, the clipped critical height c^* is

$$c^* = \begin{cases} 0 & \text{if } c \leq \mathcal{H}(m), \\ c - \mathcal{H}(m) & \text{if } c > \mathcal{H}(m). \end{cases}$$

These critical heights are illustrated in Figure 4.

Our first result gives the asymptotic order of the spectral gap $\lambda_2(-M_{\epsilon,c}^f)$ in terms of c^* in the low-temperature regime, which will be proven to be essential in obtaining a convergence result for simulated annealing.

Theorem 3. Assume that f and $\min \mathcal{H} \leq c \leq \max \mathcal{H}$ satisfy Assumption 1, and in addition for all small enough $z > 0$ we have $f(z) \geq z$. Then there exist positive constants C_2, C_3, C_4 , depending on the state space \mathcal{X} and the proposal generator Q but not on the temperature ϵ , and a subexponential function

$$C_1(\epsilon) := \begin{cases} \frac{1}{C_2} \left(1 + \frac{1}{\epsilon} (\max \mathcal{H} - c) \right) \exp \left\{ \frac{1}{f(\delta)} (\max \mathcal{H} - \min \mathcal{H}) \right\} & \text{if } c < \max \mathcal{H}, \\ \frac{1}{C_2} & \text{if } c = \max \mathcal{H}, \end{cases}$$

where $\delta := \min_{x; \mathcal{H}(x) > c} \{ \mathcal{H}(x) - c \}$, such that

$$C_1^{-1}(\epsilon) e^{-\frac{1}{\epsilon} c^*} \leq C_2 e^{-H_{\epsilon,c}^f} \leq \lambda_2(-M_{\epsilon,c}^f) \leq C_3 e^{-H_{\epsilon,c}^f} \leq C_4 e^{-\frac{1}{\epsilon} c^*},$$

where $H_{\epsilon,c}^f$ is introduced in (4.1) and c^* is defined in (4.3). This leads to

$$\lim_{\epsilon \rightarrow 0} \epsilon \log \lambda_2(-M_{\epsilon,c}^f) = -c^*.$$

As a corollary of Theorem 3, using the asymptotics of the spectral gap, we derive similar asymptotics for the mixing time and tunnelling time on the modified landscape.

Corollary 1. (Asymptotics of mixing and tunnelling times in the low-temperature regime.) *Assume the same setting as in Theorem 3. Let $S_{\min} := \arg \min \mathcal{H}(x)$ be the set of global minima of \mathcal{H} , let $\eta \in S_{\min}$, and assume that σ, η attain H^0 such that $H^0 = G^0(\sigma, \eta) - \mathcal{H}(\sigma)$. Then the following statements hold:*

1.

$$\lim_{\epsilon \rightarrow 0} \epsilon \log t_{\text{mix}}(M_{\epsilon,c}^f, 1/4) = c^*.$$

2.

$$\lim_{\epsilon \rightarrow 0} \epsilon \log \mathbb{E}_\sigma(\tau_\eta^f) = c^* \leq H^0 = \lim_{\epsilon \rightarrow 0} \epsilon \log \mathbb{E}_\sigma(\tau_\eta^0).$$

In particular, when $c = \mathcal{H}_{\min}$, we have subexponential tunnelling time, as $\lim_{\epsilon \rightarrow 0} \epsilon \log \mathbb{E}_\sigma(\tau_\eta^f) = 0 = c^*$.

Note that in the case where both $\sigma, \eta \in S_{\min}$ with initial state σ , it is a reasonable choice to pick the parameter $c = \mathcal{H}(X^f(0)) = \mathcal{H}(\sigma) = \mathcal{H}_{\min}$, and in this setting we have subexponential tunnelling time on the modified landscape. For instance, in applications we may know about a global minimizer σ , and by setting $c = \mathcal{H}(\sigma)$ we can search for other possible global minimizers owing to the subexponential tunnelling in the low-temperature regime. For a concrete example, in the Widom–Rowlinson model with $m \in \mathbb{N}$ particle types, S_{\min} is precisely the set of configurations in which all sites are occupied by particles of the same type, and hence both S_{\min} and \mathcal{H}_{\min} are known in this model. In this direction, we refer interested readers to Nardi and Zocca [9] and Zocca [10] for work on the energy landscape analysis of various statistical physics models.

To prove the convergence result for simulated annealing with landscape modification, as our target function $\mathcal{H}_{\epsilon,c}^f$ depends on time through the cooling schedule, we are in the setting of simulated annealing with time-dependent energy function as in Löwe [12]. We first present the following auxiliary lemma, where we verify various assumptions from Löwe [12] in our setting. We have included the statement of the lemma in this section, rather than with the proofs later on, since it will be helpful for understanding the convergence result in Theorem 4 below.

Lemma 1. *Assume the same setting as in Theorem 3. Let $M := \max \mathcal{H} - \min \mathcal{H}$, $\beta_t := 1/\epsilon_t$, and assume the cooling schedule is*

$$\epsilon_t = \frac{c^* + \epsilon}{\ln(t + 1)}$$

for ϵ small enough so that $M + \max \mathcal{H} - c > \epsilon > 0$ and $t \geq 0$. Then we have the following:

1. For all $x \in \mathcal{X}$ and all $t \geq 0$,

$$0 \leq \epsilon_t \mathcal{H}_{\epsilon_t, c}^f(x) \leq M.$$

2. For all $x \in \mathcal{X}$,

$$\left| \frac{\partial}{\partial t} \epsilon_t \mathcal{H}_{\epsilon_t, c}^f(x) \right| \leq \frac{2M}{(\ln(1+t))(1+t)}.$$

3. Let $R_t := \sup_x \frac{\partial}{\partial t} \epsilon_t \mathcal{H}_{\epsilon_t, c}^f(x)$ and $B := 6M/(c^* + \epsilon)$. For all $t \geq 0$,

$$\beta'_t M + \beta_t R_t \leq \frac{3M}{(c^* + \epsilon)(1+t)} = \frac{B}{2(1+t)}.$$

4. Let

$$p := \frac{2M}{M + \max \mathcal{H} - \epsilon - c} > 2$$

and

$$A := \begin{cases} \frac{1}{C_2(\min_x \mu(x))^{(p-2)/p}} \exp \left\{ \frac{1}{f(\delta)} (\max \mathcal{H} - \min \mathcal{H}) \right\} & \text{if } c < \max \mathcal{H}, \\ \frac{1}{C_2(\min_x \mu(x))^{(p-2)/p}} & \text{if } c = \max \mathcal{H}, \end{cases}$$

where C_2, δ are as in Theorem 3, and we recall that μ is the stationary measure of the proposal generator Q . For $g \in \ell^p(\pi_{\epsilon_t, c}^f)$, we have

$$\|g - \pi_{\epsilon_t, c}^f(g)\|_{\ell^p(\pi_{\epsilon_t, c}^f)}^2 \leq A(1+t) \langle -M_{\epsilon_t, c}^f g, g \rangle_{\pi_{\epsilon_t, c}^f}.$$

Remark 2. Items 1 and 2 correspond respectively to Equations (2.11) and (2.12) in Löwe [12], while the lower bound of the spectral gap in Theorem 3 verifies Equation (2.13) in Löwe [12]. Assumptions (A1) and (A2) in Löwe [12] are checked in Items 3 and 4, respectively.

With the above lemma and the notation introduced there, we are ready to give one of the main results of this paper, which concerns the large-time convergence of discrete simulated annealing with landscape modification. The gain from landscape modification in simulated annealing can be seen by operating a possibly faster logarithmic cooling schedule with clipped critical height c^* , while in classical simulated annealing the critical height is H^0 . The possible benefit thus depends on the tuning of c , since $c^* \leq c - \mathcal{H}_{\min}$. This result is analogous to the result obtained in Choi [4] for improved kinetic simulated annealing. A similar improvement of a logarithmic cooling schedule by means of reduction in critical height can be found in the infinite swapping algorithm [17].

Theorem 4. Assume the same setting as in Theorem 3. Let A, B, p be the quantities introduced in Lemma 1. Define

$$\bar{\epsilon} := \frac{p-2}{p}, \quad K := \frac{4(1+2AB)}{1 - \exp\{-\frac{1}{2A} - B\}}, \quad S_{\min} := \arg \min \mathcal{H}, \quad \underline{d} := \min_{x: \mathcal{H}(x) \neq \mathcal{H}_{\min}} \mathcal{H}(x).$$

Consider a cooling schedule of the following form: for any $\epsilon > 0$ as in Lemma 1,

$$\epsilon_t = \frac{c^* + \epsilon}{\ln(t + 1)}.$$

Under this cooling schedule, for any $x \in \mathcal{X} \setminus S_{\min}$ and $t \geq e^{1/\epsilon} - 1$, we have

$$\mathbb{P}_x \left(\tau_{S_{\min}}^f > t \right) \leq (1 + K^{\frac{1}{2\epsilon}}) \sqrt{\pi_{\epsilon_t, c}^f(\mathcal{X} \setminus S_{\min}) + \pi_{\epsilon_t, c}^f(\mathcal{X} \setminus S_{\min})} \rightarrow 0 \quad \text{as } t \rightarrow \infty.$$

Note that

$$\pi_{\epsilon_t, c}^f(\mathcal{X} \setminus S_{\min}) \leq \begin{cases} \frac{1}{\mu(S_{\min})} e^{-\frac{1}{\epsilon_t}(\underline{d} - \mathcal{H}_{\min})} & \text{if } c \geq \underline{d}, \\ \frac{1}{\mu(S_{\min})} \exp \left\{ - \int_c^{\underline{d}} \frac{1}{f(u-c) + \epsilon_t} du \right\} & \text{if } \mathcal{H}_{\min} \leq c < \underline{d}. \end{cases}$$

Remark 3. (On tuning the threshold parameter c .) There are various ways to tune the parameter c for improved convergence. In Choi [4], we proposed to use the running minimum generated by the algorithm to tune c . Note that in the setting of the CW model, the parameter c can be tuned as explained earlier, in the second paragraph below Theorem 1.

4.1. Numerical illustrations

Before discussing the proofs of the main results above, we illustrate the convergence performance of simulated annealing with landscape modification, which we call *improved simulated annealing* (ISA), and compare it to the performance of the classical simulated annealing algorithm (SA) for the travelling salesman problem (TSP). We first state the parameters that we used to generate the numerical results:

- **TSP and its objective function.** We select 50 nodes uniformly at random on the grid $[0, 100] \times [0, 100]$. The objective is to find a configuration that minimizes the total Euclidean distance with the same starting and ending point. Each node can only be visited once.
- **Initial configuration.** Both ISA and SA have the same initialization. They are initialized using the output of the nearest-neighbour algorithm: a node is randomly chosen as the starting point, which is then connected to the closest unvisited node. This is repeated until every node has been visited, and subsequently the last node is connected back to the starting node.
- **Proposal chain.** Both ISA and SA share the same proposal chain: at each step, a proposal move is generated using the 2-OPT algorithm [18].
- **Acceptance–rejection mechanism.** In SA, the proposed move is accepted with probability

$$\min \left\{ 1, e^{\beta(\mathcal{H}(x) - \mathcal{H}(y))} \right\},$$

while in ISA, the acceptance probability is computed as in Section 2.1. In other words, we use a linear f in ISA. Both SA and ISA have the same source of randomness.

TABLE 1. Summary statistics of IP on 1000 random TSP instances.

Sample mean	1.87%
Sample median	1.47%
Sample maximum	11.35%
Sample minimum	-9.21%
Number of IP ≥ 0	798
Number of IP < 0	202

- **Cooling schedule.** Both ISA and SA use the same logarithmic cooling schedule, of the form

$$\epsilon_t = \frac{\sqrt{50}}{\ln(t+1)}.$$

- **Choice of c in ISA.** In this experiment, if we denote the proposal configuration at time t by y_t , we set $c = c_t$ to be

$$c_t = \mathcal{H}(y_t) - 5.$$

This tuning strategy has been discussed in Section 2.5.

- **Number of iterations.** We run both ISA and SA for 100,000 iterations.

We generate 1000 random TSP instances. For each instance we compute what we call the improvement percentage (IP) of ISA over SA, defined by

$$\text{IP} := 100 \frac{\min_{t \in [0, 100,000]} \mathcal{H}(X^0(t)) - \min_{t \in [0, 100,000]} \mathcal{H}(X^f(t))}{\min_{t \in [0, 100,000]} \mathcal{H}(X^0(t))}.$$

The summary statistics of IP are provided in Table 1, while its histogram over these 1000 instances can be found in Figure 5. The code for reproducing these results can be found at <https://github.com/mchchoi/Improved-discrete-simulated-annealing>.

The summary statistics in Table 1 and the histogram in Figure 5 offer empirical evidence for using ISA over SA: out of the 1000 TSP instances, there are 798 instances in which the IP is non-negative. The sample mean of the IP is approximately 1.87%, while its sample median is 1.47%.

Next, in Figure 6, we investigate the difference between SA and ISA in a particular instance. We see that SA (the blue curve) is stuck at a local minimum, while ISA (the orange curve) is able to escape the local minimum, owing to the increased acceptance probability compared with SA, and it reaches regions where the objective value is smaller than that of SA.

The rest of this section is devoted to the proofs of Theorem 3, Corollary 1, Lemma 1, and Theorem 4.

4.2 Proof of Theorem 3

First, using the classical result of Holley and Stroock ([19], Theorem 2.1), it is immediate that

$$C_2 e^{-H_{\epsilon,c}^f} \leq \lambda_2(-M_{\epsilon,c}^f) \leq C_3 e^{-H_{\epsilon,c}^f}.$$

For any arbitrary $x_1, x_2 \in \{\mathcal{H}(x_1) \geq \mathcal{H}(x_2)\}$, we deduce the following upper bound:

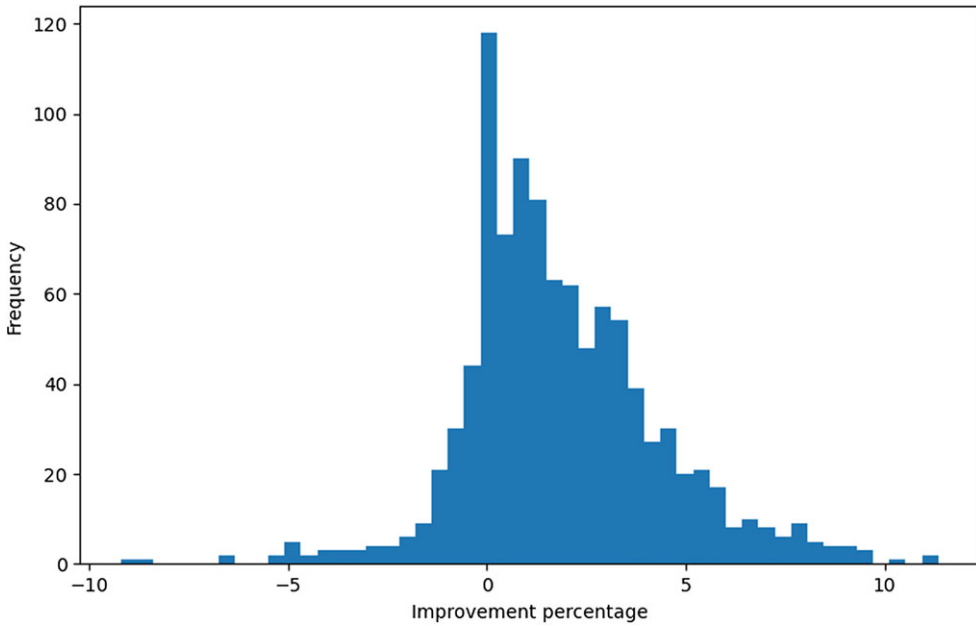


FIGURE 5. Histogram showing IP of ISA over SA on 1000 randomly generated TSP instances.

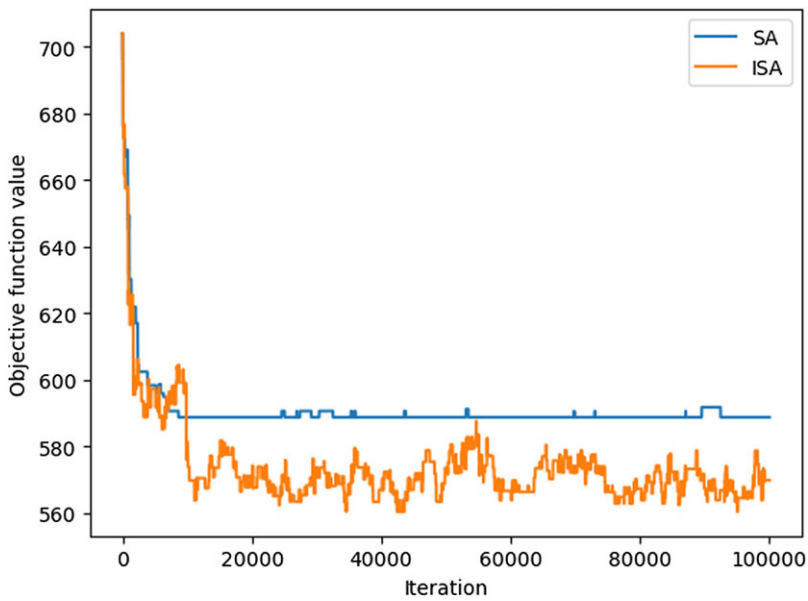


FIGURE 6. TSP objective value against iteration of ISA and SA.

$$\begin{aligned}
 & \mathcal{H}_{\epsilon,c}^f(x_1) - \mathcal{H}_{\epsilon,c}^f(x_2) \\
 &= \int_{\mathcal{H}(x_2)}^{\mathcal{H}(x_1)} \frac{1}{f((u-c)_+) + \epsilon} du \\
 &= \begin{cases} \frac{1}{\epsilon}(\mathcal{H}(x_1) - \mathcal{H}(x_2)) & \text{if } c \geq \mathcal{H}(x_1) > \mathcal{H}(x_2), \\ \frac{1}{\epsilon}(c - \mathcal{H}(x_2)) + \int_c^{\mathcal{H}(x_1)} \frac{1}{f(u-c) + \epsilon} du & \text{if } \mathcal{H}(x_1) > c \geq \mathcal{H}(x_2), \\ \int_{\mathcal{H}(x_2)}^{\mathcal{H}(x_1)} \frac{1}{f(u-c) + \epsilon} du & \text{if } \mathcal{H}(x_1) > \mathcal{H}(x_2) > c, \end{cases} \\
 &\leq \begin{cases} \frac{1}{\epsilon}(\mathcal{H}(x_1) \wedge c - \mathcal{H}(x_2) \wedge c) & \text{if } c \geq \mathcal{H}(x_1) > \mathcal{H}(x_2), \\ \frac{1}{\epsilon}(\mathcal{H}(x_1) \wedge c - \mathcal{H}(x_2) \wedge c) + \ln\left(1 + \frac{1}{\epsilon}(\max \mathcal{H} - c)\right) & \text{if } \mathcal{H}(x_1) > c \geq \mathcal{H}(x_2), \\ \frac{1}{f(\delta)}(\max \mathcal{H} - \min \mathcal{H}) & \text{if } \mathcal{H}(x_1) > \mathcal{H}(x_2) > c. \end{cases}
 \end{aligned}$$

As a result, $C_2^{-1} e^{\mathcal{H}_{\epsilon,c}^f} \leq e^{\frac{1}{\epsilon}c^*} C_1(\epsilon)$. On the other hand, we have the following lower bound:

$$\mathcal{H}_{\epsilon,c}^f(x_1) - \mathcal{H}_{\epsilon,c}^f(x_2) \geq \begin{cases} \frac{1}{\epsilon}(\mathcal{H}(x_1) \wedge c - \mathcal{H}(x_2) \wedge c) & \text{if } c \geq \mathcal{H}(x_1) > \mathcal{H}(x_2), \\ \frac{1}{\epsilon}(\mathcal{H}(x_1) \wedge c - \mathcal{H}(x_2) \wedge c) & \text{if } \mathcal{H}(x_1) > c \geq \mathcal{H}(x_2), \\ \frac{1}{f(\mathcal{H}(x_1) - c)}(\mathcal{H}(x_1) - \mathcal{H}(x_2)) & \text{if } \mathcal{H}(x_1) > \mathcal{H}(x_2) > c, \end{cases}$$

and hence $e^{\mathcal{H}_{\epsilon,c}^f} \geq e^{\frac{1}{\epsilon}c^*} C_4^{-1}$.

4.3. Proof of Corollary 1

We first prove Item 1. For a continuous-time reversible Markov chain, by Levin and Peres ([20], Theorems 12.5 and 20.6), we can bound the total-variation mixing time by relaxation time via

$$\frac{1}{\lambda_2(-M_{\epsilon,c}^f)} \log 2 \leq t_{mix}(M_{\epsilon,c}^f, 1/4) \leq \frac{1}{\lambda_2(-M_{\epsilon,c}^f)} \log\left(\frac{4}{\pi_{min}^f}\right),$$

where

$$\pi_{min}^f := \min_x \pi_{\epsilon,c}^f(x) = \frac{\mu(x^*)}{Z_{\epsilon,c}^f}$$

for some $x^* \in S_{min}$, and

$$Z_{\epsilon,c}^f := \sum_{x \in \mathcal{X}} e^{-\mathcal{H}_{\epsilon,c}^f(x)} \mu(x)$$

is the normalization constant. Note that since $\ln Z_{\epsilon,c}^f \rightarrow \ln \mu(S_{\min})$, we have

$$\lim_{\epsilon \rightarrow 0} \epsilon \ln Z_{\epsilon,c}^f = 0.$$

Item 1 follows from collecting the above results together with Theorem 3.

Next, we prove Item 2. First, using the random target lemma ([15], Section 4.2), we have

$$\pi_{\epsilon,c}^f(\eta) \mathbb{E}_\sigma(\tau_\eta^f) \leq \sum_{x \in \mathcal{X}} \pi_{\epsilon,c}^f(x) \mathbb{E}_\sigma(\tau_x^f) \leq (|\mathcal{X}| - 1) \frac{1}{\lambda_2(-M_{\epsilon,c}^f)}.$$

Since $\mathcal{H}(\eta) = 0$ and $Z_{\epsilon,c}^f \leq 1$, upon rearranging and using Theorem 2 we obtain

$$\limsup_{\epsilon \rightarrow 0} \epsilon \log \mathbb{E}_\sigma(\tau_\eta^f) \leq \lim_{\epsilon \rightarrow 0} \epsilon \log \frac{1}{\lambda_2(-M_{\epsilon,c}^f)} = c^*.$$

Define the equilibrium potential and capacity of the pair (σ, η) as in Bovier and den Hollander ([14], Chapter 7.2) to be respectively

$$h_{\sigma,\eta}^f(x) := \mathbb{P}_x(\tau_\sigma^f < \tau_\eta^f),$$

$$\text{cap}^{M_{\epsilon,c}^f}(\sigma, \eta) := \inf_{f: f|_A=1, f|_B=0} \langle -M_{\epsilon,c}^f, f \rangle_{\pi_{\epsilon,c}^f} = \langle -M_{\epsilon,c}^f, h_{\sigma,\eta}^f \rangle_{\pi_{\epsilon,c}^f}.$$

If we prove that

$$\text{cap}^{M_{\epsilon,c}^f}(\sigma, \eta) \leq \frac{1}{Z_{\epsilon,c}^f} \left(\sum_{x,y} \mu(x) Q(x, y) \right) e^{-\frac{1}{\epsilon} G^0(\sigma, \eta)}, \tag{4.4}$$

then the mean hitting time formula combined with the equilibrium potential and capacity leads to

$$\begin{aligned} \mathbb{E}_\sigma(\tau_\eta^f) &= \frac{1}{\text{cap}^{M_{\epsilon,c}^f}(\sigma, \eta)} \sum_{y \in \mathcal{X}} \pi_{\epsilon,c}^f(y) h_{\sigma,\eta}^f(y) \\ &\geq \frac{1}{\text{cap}^{M_{\epsilon,c}^f}(\sigma, \eta)} \pi_{\epsilon,c}^f(\sigma) h_{\sigma,\eta}^f(\sigma) \geq \frac{\mu(\sigma)}{\sum_{x,y} \mu(x) Q(x, y)} e^{\frac{1}{\epsilon} c^*}, \end{aligned}$$

and the desired result follows, since

$$\liminf_{\epsilon \rightarrow 0} \epsilon \log \mathbb{E}_\sigma(\tau_\eta^f) \geq c^*.$$

It therefore remains to prove (4.4). Define

$$\Phi(\sigma, \eta) := \{x \in \mathcal{X}; G^0(x, \sigma) \leq G^0(x, \eta)\}.$$

Let $\mathbf{1}_A$ denote the indicator function of the set A ; the Dirichlet principle of capacity gives

$$\begin{aligned} \text{cap}^{M_{\epsilon,c}^f}(\sigma, \eta) &\leq \langle -M_{\epsilon,c}^f \mathbf{1}_{\Phi(\sigma, \eta)}, \mathbf{1}_{\Phi(\sigma, \eta)} \rangle_{\pi_{\epsilon,c}^f} = \frac{1}{Z_{\epsilon,c}^f} \sum_{x \in \Phi(\sigma, \eta), y \notin \Phi(\sigma, \eta)} e^{-\frac{1}{\epsilon} (H(x) \vee H(y))} \mu(x) Q(x, y) \\ &\leq \frac{1}{Z_{\epsilon,c}^f} \left(\sum_{x,y} \mu(x) Q(x, y) \right) e^{-\frac{1}{\epsilon} G^0(\sigma, \eta)}, \end{aligned}$$

where in the last inequality we use the fact that $G^0(\sigma, \eta)$ is the lowest possible highest elevation between σ and η .

4.4. Proof of Lemma 1

We first prove Item 1. The lower bound is immediate, while the upper bound can be deduced via

$$\epsilon_t \mathcal{H}_{\epsilon_t, c}^f \leq \epsilon_t \int_{\mathcal{H}_{\min}}^{\mathcal{H}(x)} \frac{1}{\epsilon_t} du \leq M.$$

Next, we prove Item 2. We consider

$$\begin{aligned} \frac{\partial}{\partial t} \epsilon_t \mathcal{H}_{\epsilon_t, c}^f(x) &= \mathcal{H}_{\epsilon_t, c}^f(x) \left(\frac{\partial}{\partial t} \epsilon_t \right) + \epsilon_t \frac{\partial}{\partial t} \mathcal{H}_{\epsilon_t, c}^f(x) \\ &= \mathcal{H}_{\epsilon_t, c}^f(x) \frac{-c^* - \epsilon}{(\ln(t+1))^2} \frac{1}{t+1} + \epsilon_t \epsilon_t' \int_{\mathcal{H}_{\min}}^{\mathcal{H}(x)} -\frac{1}{(f((u-c)_+) + \epsilon_t)^2} du. \end{aligned}$$

This leads to

$$\left| \frac{\partial}{\partial t} \epsilon_t \mathcal{H}_{\epsilon_t, c}^f(x) \right| \leq \frac{2M}{(\ln(t+1))(t+1)}.$$

Thirdly, we prove Item 3. Using Item 2, we calculate that

$$\beta_t' M + \beta_t R_t \leq \frac{M}{(c^* + \epsilon)(t+1)} + \frac{2M}{(c^* + \epsilon)(1+t)} = \frac{3M}{(c^* + \epsilon)(1+t)}.$$

Finally, we prove Item 4. Following exactly the same calculation as in the proof of Löwe ([12], Lemma 3.5), we see that

$$\begin{aligned} \left\| g - \pi_{\epsilon_t, c}^f(g) \right\|_{\ell^p(\pi_{\epsilon_t, c}^f)}^2 &\leq A (1 + \beta_t(\max \mathcal{H} - c)) e^{\beta_t(c^* + M \frac{p-2}{p})} \langle -M_{\epsilon_t, c}^f g, g \rangle_{\pi_{\epsilon_t, c}^f} \\ &\leq A e^{\beta_t(c^* + M \frac{p-2}{p} + \max \mathcal{H} - c)} \langle -M_{\epsilon_t, c}^f g, g \rangle_{\pi_{\epsilon_t, c}^f}. \end{aligned}$$

The desired result follows if we let

$$\epsilon = \frac{M(p-2)}{p} + \max \mathcal{H} - c,$$

so that

$$p = \frac{2M}{M + \max \mathcal{H} - \epsilon - c}.$$

4.5 Proof of Theorem 4

We would like to invoke the results in Löwe [12] for a time-dependent target function in simulated annealing. In Lemma 1, Items 1, 2, 3, and 4, we verify that Equations (11) and (12) and Assumptions (A1) and (A2), respectively, from Löwe [12] hold. Consequently, if we let

$h_t(y) := \mathbb{P}_x(X_{\epsilon_t, c}^f(t) = y) / \pi_{\epsilon_t, c}^f(y)$, then according to Holley and Stroock ([19], Lemma 1.7), its ℓ^2 norm is bounded by

$$\|h_t\|_{\ell^2(\pi_{\epsilon_t, c}^f)} \leq 1 + K^{\frac{1}{2\bar{\epsilon}}},$$

for $t \geq e^{1/\bar{\epsilon}} - 1$. The desired result follows from exactly the same argument as in Löwe ([12], Theorem 3.8).

Now we calculate $\pi_{\epsilon_t, c}^f(\mathcal{X} \setminus S_{\min})$. For $c \geq \underline{d}$, we compute that

$$\begin{aligned} \pi_{\epsilon_t, c}^f(\mathcal{X} \setminus S_{\min}) &= \frac{\sum_{x \in \mathcal{X} \setminus S_{\min}} e^{-\mathcal{H}_{\epsilon_t, c}^f(x)} \mu(x)}{\sum_{x \in \mathcal{X}} e^{-\mathcal{H}_{\epsilon_t, c}^f(x)} \mu(x)} \\ &\leq \frac{\sum_{x: \mathcal{H}(x) \geq \underline{d}} \mu(x)}{\sum_{x: \mathcal{H}(x) \leq \underline{d}} \exp\left\{\int_{\mathcal{H}(x)}^{\underline{d}} \frac{1}{\epsilon_t} du\right\} \mu(x)} \\ &\leq e^{-\frac{1}{\epsilon_t}(\underline{d} - \mathcal{H}_{\min})} \frac{1}{\mu(S_{\min})}. \end{aligned}$$

On the other hand, for $\mathcal{H}_{\min} \leq c < \underline{d}$,

$$\begin{aligned} \pi_{\epsilon_t, c}^f(\mathcal{X} \setminus S_{\min}) &= \frac{\sum_{x \in \mathcal{X} \setminus S_{\min}} e^{-\mathcal{H}_{\epsilon_t, c}^f(x)} \mu(x)}{\sum_{x \in \mathcal{X}} e^{-\mathcal{H}_{\epsilon_t, c}^f(x)} \mu(x)} \\ &\leq \frac{1}{\mu(S_{\min})} \exp\left\{-\int_{\mathcal{H}_{\min}}^{\underline{d}} \frac{1}{f((u-c)_+) + \epsilon_t} du\right\} \\ &\leq \frac{1}{\mu(S_{\min})} \exp\left\{-\int_c^{\underline{d}} \frac{1}{f(u-c) + \epsilon_t} du\right\}. \end{aligned}$$

Acknowledgements

We thank Laurent Miclo for pointers to the work of Olivier Catoni, and we thank the two reviewers for careful reading and constructive feedback.

Funding information

The author acknowledges financial support from the startup grant of the National University of Singapore, Yale-NUS College, and a Singapore MoE Tier 1 grant entitled ‘MAPLE’.

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

References

- [1] CATONI, O. (1996). Metropolis, simulated annealing, and iterated energy transformation algorithms: theory and experiments. *J. Complexity* **12**, 595–623.
- [2] CATONI, O. (1998). The energy transformation method for the Metropolis algorithm compared with simulated annealing. *Prob. Theory Relat. Fields* **110**, 69–89.

- [3] WANG, Y., WU, S. AND ZOU, J. (2016). Quantum annealing with Markov chain Monte Carlo simulations and D-wave quantum computers. *Statist. Sci.* **31**, 362–398.
- [4] CHOI, M. C. (2020). *On the convergence of an improved and adaptive kinetic simulated annealing*. Preprint. Available at <https://arxiv.org/abs/2009.00195>.
- [5] FANG, H., QIAN, M. AND GONG, G. (1997). An improved annealing method and its large-time behavior. *Stoch. Process. Appl.* **71**, 55–74.
- [6] MONMARCHÉ, P. (2018). Hypocoercivity in metastable settings and kinetic simulated annealing. *Prob. Theory Relat. Fields* **172**, 1215–1248.
- [7] ZHANG, J. AND CHOI, M. C. (2021). Improved simulated annealing for sampling from multimodal distributions. Working paper.
- [8] DEUSCHEL, J.-D. AND MAZZA, C. (1994). L^2 convergence of time nonhomogeneous Markov processes: I. Spectral estimates. *Ann. Appl. Prob.* **4**, 1012–1056.
- [9] NARDI, F. R. AND ZOCCA, A. (2019). Tunneling behavior of Ising and Potts models in the low-temperature regime. *Stoch. Process. Appl.* **129**, 4556–4575.
- [10] ZOCCA, A. (2018). Low-temperature behavior of the multicomponent Widom–Rowlison model on finite square lattices. *J. Statist. Phys.* **171**, 1–37.
- [11] DEL MORAL, P. AND MICLO, L. (1999). On the convergence and applications of generalized simulated annealing. *SIAM J. Control Optimization* **37**, 1222–1250.
- [12] LÖWE, M. (1996). Simulated annealing with time-dependent energy function via Sobolev inequalities. *Stoch. Process. Appl.* **63**, 221–233.
- [13] FRIGERIO, A. AND GRILLO, G. (1993). Simulated annealing with time-dependent energy function. *Math. Z.* **213**, 97–116.
- [14] BOVIER, A. AND DEN HOLLANDER, F. (2015). *Metastability: A Potential-Theoretic Approach*. Springer, Cham.
- [15] ALDOUS, D. AND FILL, J. A. (2002). *Reversible Markov Chains and Random Walks on Graphs*. Unfinished monograph, recompiled 2014. Available at <http://www.stat.berkeley.edu/aldous/RWG/book.html>.
- [16] MATHIEU, P. AND PICCO, P. (1998). Metastability and convergence to equilibrium for the random field Curie–Weiss model. *J. Statist. Phys.* **91**, 679–732.
- [17] MENZ, G., SCHLICHTING, A., TANG, W. AND WU, T. (2022). Ergodicity of the infinite swapping algorithm at low temperature. *Stoch. Process. Appl.* **151**, 519–552.
- [18] CROES, G. A. (1958). A method for solving traveling-salesman problems. *Operat. Res.* **6**, 791–812.
- [19] HOLLEY, R. AND STROOCK, D. (1988). Simulated annealing via Sobolev inequalities. *Commun. Math. Phys.* **115**, 553–569.
- [20] LEVIN, D. A. AND PERES, Y. (2017). *Markov Chains and Mixing Times*. American Mathematical Society, Providence, RI.