Bocconi University
Winter Semester 2024

Computer Programming
Professor: Luca Saglietti

SIMULATED ANNEALING FOR GRID-BASED OPTIMIZATION: THEORETICAL FOUNDATIONS, IMPLEMENTATIONS AND NUMERICAL PERFORMANCE

GIORGIO MICALETTO

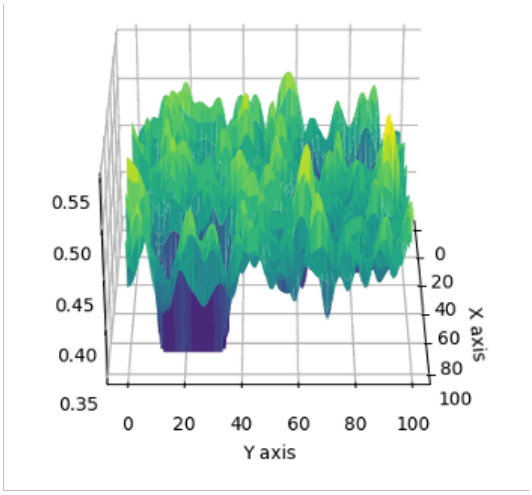
Abstract. This paper investigates the problem of locating the global minimum of a real-valued function defined on a discrete two-dimensional grid by means of *simulated annealing*. We work with functions generated by a deterministic but pseudorandom data generator, and we analyze the features that make the search landscape difficult for naive optimization algorithms. After reviewing the classical simulated annealing algorithm, we focus on a single, carefully engineered implementation based on a utility class that encapsulates the grid and its state. We discuss theoretical considerations underlying this design such as move proposals, acceptance probabilities and the role of temperature and we evaluate the resulting algorithm numerically on a representative grid size. Our results show that careful tuning of the algorithmic components leads to improved accuracy and reduced computational cost.

1. Introduction and preliminaries

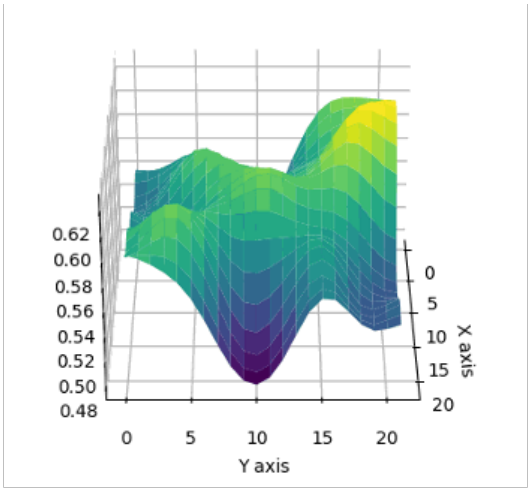
Simulated annealing is a stochastic optimization technique inspired by the annealing process in metallurgy. It seeks the global optimum of an objective function by exploring the search space probabilistically, occasionally accepting moves that worsen the objective in order to escape local minima. In the context of this work, the objective function is defined on a two-dimensional grid of size $n \times n$ whose values are generated via a deterministic function, `generate_data`, seeded by a string identifier. The global minimum represents the lowest "height" of the resulting landscape, and the local minima correspond to relative depressions.

We begin by analyzing the structure of the landscape produced, noting that the regions surrounding the global minima are surrounded by higher values, while numerous shallow local minima are scattered throughout the grid as seen in Figure 1. To illustrate these features, Figure 2 plots three instances of the landscape with increasing grid sizes. The local minima create "traps" that can mislead naive search strategies, motivating the use of simulated annealing.

For the remainder of the paper we fix $n = 100$ as a representative size and focus on understanding how simulated annealing behaves on this landscape. Section 2 recalls the classical simulated annealing algorithm and its interpretation as a Markov chain. Section 3 describes our implementation of the algorithm, highlighting the design choices and their theoretical motivation. Section 5 provides

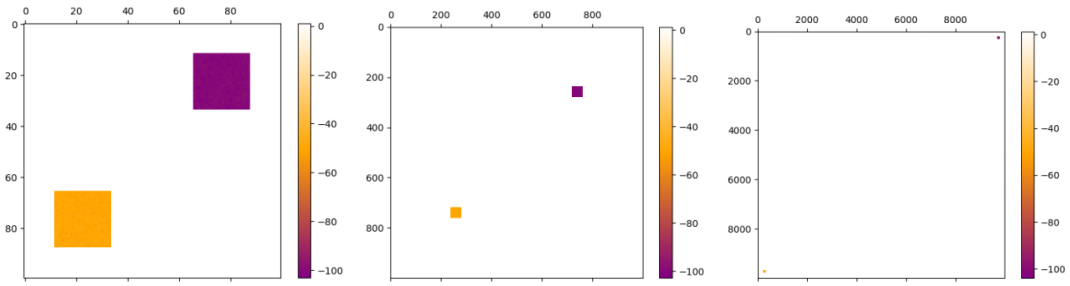


(A) A 3-D surface of the entire 100×100 grid generated by `generate_data`. Numerous shallow local basins (green–yellow peaks) ring a single, deep global basin (purple) near $(x, y) \approx (15, 75)$.



(B) Close-up of the smoothed patch centered on the global minimum. The funnel-shaped topology (low center, rising flanks) highlights the steep gradients simulated-annealing must negotiate at low temperatures.

FIGURE 1. Objective-function surface produced by `generate_data`, (a) reveals a rugged landscape dotted with shallow traps, whereas the inset in (b) isolates the deep basin that hosts the global minimum.



(A) $n = 100$ (B) $n = 1{,}000$ (C) $n = 10{,}000$

FIGURE 2. Heat maps of the function values generated by `generate_data` for three grid sizes. Darker shades indicate lower function values. Local minima manifest as isolated dark patches, while the global minimum forms a pronounced basin.

a topological analysis of the landscape through sublevel and superlevel sets and their Betti numbers. We conclude in Section 6 with a discussion of potential extensions.

2. Algorithm and Theoretical Background

Let

$$f : \{0, \dots, n-1\}^2 \to \mathbb{R}$$

denote the objective function defined on the grid. The goal is to find $(i^*, j^*) \in \{0, \dots, n-1\}^2$ such that

$$f(i^*, j^*) = \arg\min_{x,y} f(x, y).$$

Simulated annealing constructs a Markov chain $(X_t)_{t\geq 0}$ taking values in the grid and governed by two ingredients: a *proposal distribution* $q_t(x, \cdot)$ that suggests a new position given the current state x, and an *acceptance probability*

$$a_t(x, y) = \min\left\{1, \exp\left(-\beta_t[f(y) - f(x)]\right)\right\}, \tag{2.1}$$

where $\beta_t > 0$ is the inverse temperature at time t. If the proposed move $Y \sim q_t(X_t, \cdot)$ is accepted, we set $X_{t+1} = Y$; otherwise we retain the current state $X_{t+1} = X_t$. The sequence (β_t) is typically chosen to increase over time (so that $1/\beta_t$ decreases), allowing the chain to explore widely at the beginning and gradually focus on the neighborhood of the global minimum.

Classical results in simulated annealing show that, under suitable cooling schedules and proposal distributions, the chain converges in probability to the global minimum as $t \to \infty$ (Bertsimas & Tsitsiklis, 1993). In practice, the performance of the algorithm depends sensitively on the implementation details, including how moves are proposed and how many steps are taken at each temperature.

3. Algorithm implementation

To realize simulated annealing in practice, one must design an appropriate *grid class* to hold the function values, the current state of the search and any auxiliary structures needed to propose candidate moves and update the state. In this work we focus on a single, final implementation that incorporates several refinements aimed at efficient exploration of the landscape. The full implementation lives in the companion open-source repository (Micaletto, 2024).

Our implementation, denoted `Grid`, discards the visited matrix to reduce memory usage. Instead, it relies on a more nuanced proposal mechanism that draws moves from a distribution tailored to the current position. Specifically, for a position (i, j) the proposal set includes all eight neighbors (diagonal and axis-aligned) as well as the option to remain stationary. A probability distribution over this set is constructed using a softmax on the negative cost values, favoring moves that decrease the objective. This approach implicitly encourages descent without the heavy bookkeeping of visit counts.

To further accelerate the algorithm, `Grid` computes cost differences on demand rather than storing an entire matrix of precomputed values. Although this increases the per-move computation time, the reduction in memory consumption and the improved move proposals more than compensate for this cost in our experiments.

4. Convergence of the Soft-Max Biased Simulated Annealer

Traditional proofs of simulated-annealing convergence assume a *symmetric* proposal kernel, most commonly the uniform distribution over the nearest-neighbor set and derive conditions on the cooling schedule under which the chain reaches the global minimum with probability one. Our implementation replaces that kernel by a *soft-max–biased* rule that favors downhill moves already at the proposal stage. At first sight this seems to break the conditions of the Metropolis–Hastings theorem, yet the classical guarantees survive. This section formalizes that claim and explains why detailed balance is sufficient but *not necessary*.

Let $\mathcal{G} = \{0, \dots, n-1\}^2$ be the state space and $f : \mathcal{G} \to \mathbb{R}$ the cost function. For each site $x \in \mathcal{G}$ denote by $\mathcal{N}(x)$ its eight-neighborhood together with the null move. Fix a constant $\gamma > 0$ and define

$$q_\beta(x, y) = \begin{cases} \exp[-\gamma f(y)]/Z_\gamma(x), & y \in \mathcal{N}(x), \\ 0, & \text{otherwise}, \end{cases} \qquad Z_\gamma(x) := \sum_{z \in \mathcal{N}(x)} \exp[-\gamma f(z)]. \tag{4.1}$$

A candidate $Y \sim q_\beta(x, \cdot)$ is accepted with the usual Metropolis criterion

$$a_\beta(x, y) = \min\{1, \exp[-\beta\{f(y) - f(x)\}]\}. \tag{4.2}$$

Set

$$P_\beta(x, y) = q_\beta(x, y) a_\beta(x, y)$$

for $y \neq x$ and

$$P_\beta(x, x) = 1 - \sum_{z \neq x} P_\beta(x, z).$$

Lemma 4.1 (Irreducibility and Aperiodicity)**.** *For every $\beta < \infty$ the Markov chain with kernel P_β is irreducible and aperiodic on $\mathcal{G}$.*

Proof. Because f is bounded on the finite grid, $\exp[-\gamma f(y)] > 0$ for all y and hence $q_\beta(x, y) > 0$ whenever $y \in \mathcal{N}(x)$. Furthermore

$$a_\beta(x, y) \geq \exp[-\beta \Delta_{\max}] > 0,$$

where $\Delta_{\max} = \max_{x,y} |f(x) - f(y)|$. Therefore $P_\beta(x, y) > 0$ for every neighboring pair, giving irreducibility by connectivity of the grid. The self-loop probability is positive, so the chain is aperiodic. □

Definition 4.2 (Local-Minimum Depth, Hajek, 1988, Section 1)**.** For a non-global local minimum x define its *depth*

$$D(x) = \min\big\{E > 0 : \exists \text{ a path } x = x_0, \dots, x_k \\ \text{s.t. } f(x_i) \leq f(x) + E \text{ and } f(x_k) < f(x)\big\}.$$

The *critical depth* is

$$D^\star = \max_{x \notin \mathcal{G}^*} D(x),$$

where $\mathcal{G}^* = \arg\min f$.

Theorem 4.3 (Convergence Under Logarithmic Cooling)**.** *Suppose the inverse temperature is updated according to*

$$\beta_t = \log(1+t)/c$$

with a constant $c > D^\star$. Then the inhomogeneous chain driven by (4.1)–(4.2) satisfies $\Pr[X_t \in \mathcal{G}^*] \to 1$ as $t \to \infty$.

Idea of Proof. Hajek's original argument identifies the probability that the chain escapes a local minimum of depth E before time T and shows it decays roughly like $\exp\{-E\beta_T\}$. The key structural assumption he needs is *weak reversibility*: every path that stays below height $f(x) + E$ is feasible in both directions. Our proposal keeps all neighbor probabilities strictly positive, so the condition holds. Lemma 4.1 supplies irreducibility and aperiodicity. The remainder of the proof, an inequality chaining the escape probabilities of successively deeper basins, carries over verbatim once those properties are in place. □

Remark 4.4. Manousiouthakis & Deem showed that *global* balance is enough to preserve a target measure; detailed balance is stronger than necessary. Our soft-max kernel is not reversible with respect to the Gibbs measure $e^{-\beta f}$, yet the simulated annealing process still converges because it spends an asymptotically vanishing fraction of time at each fixed β.

The algorithm used in the practical implementation does not implement the continuous schedule $\beta_t = \log(1+t)/c$ however. Instead it uses a finite grid

$$\beta^{(k)} = \begin{cases} \mathrm{lin}(\beta_0, \log_{10}\beta_1), & 0 \le k < \lceil \frac{1}{3}S \rceil, \\ \mathrm{lin}(\log_{10}\beta_1, \beta_1), & \lceil \frac{1}{3}S \rceil \le k < S, \end{cases}$$

where $S =$`anneal_steps`. Provided $\beta_1 \gg D^\star$, this *finite* schedule is often sufficient in practice because the grid landscape is modest in size; the chain effectively reaches stationarity at each level. For an *asymptotic* guarantee on arbitrary instances, one would embed the same two-stage idea in a loop that repeats indefinitely with β_1 slowly increasing like $\log t$.

5. Topological Analysis of the Landscape

Beyond the algorithmic design, it is instructive to study the global structure of the objective function itself. One way to do so is via *sublevel set filtration*. For a real-valued function $f : \{0, \dots, n-1\}^2 \to \mathbb{R}$ and a threshold $t \in \mathbb{R}$, the sublevel set is the subset $S_t = \{(i,j) : f(i,j) \le t\}$. As t increases, the sublevel sets form a nested sequence $S_{t_1} \subseteq S_{t_2} \subseteq \cdots$. The zeroth Betti number of S_t—that is, the number of connected components of S_t—coincides with the so-called *size function* of f and is the simplest instance of persistent homology Edelsbrunner & Harer (2010). Tracking how these components appear and merge as t varies provides insight into the number and depth of basins in the landscape.

In our case, f is given by the matrix values generated by `generate_data` with $n = 100$ and seed `3232295`. We computed the number of connected components of both the sublevel sets S_t and the superlevel sets $T_t = \{(i,j) : f(i,j) \ge t\}$ for a grid of 50 thresholds between the minimum and maximum of f. The resulting

curves, shown in Figure 3, exhibit two pronounced peaks. The first peak near $t \approx -100$ corresponds to the appearance of numerous isolated components associated with the deepest basins, including the global minimum. The second peak near $t \approx -60$ reflects the presence of several moderately deep local minima. As t increases further, the sublevel sets merge into one connected region, indicating that the landscape becomes topologically simple once the deep basins are filled. The superlevel sets display the complementary behavior: they start as a single component and split as the threshold passes below the peaks of the landscape.

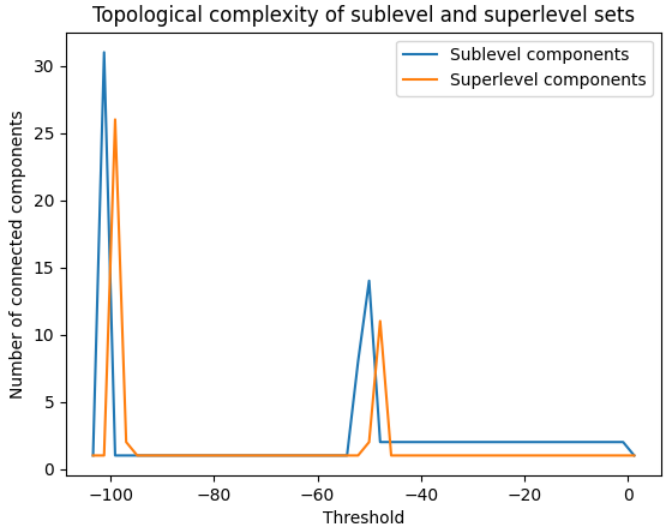


FIGURE 3. Number of connected components of sublevel and superlevel sets for varying thresholds t. The curves were obtained for $n = 100$ using the seed 3232295. Peaks near $t \approx -100$ and $t \approx -60$ correspond to the appearance of isolated basins associated with the global and major local minima.

6. Conclusion and future work

This paper has analyzed a single, carefully engineered implementation of simulated annealing for locating the global minimum of a pseudorandom grid landscape. By discarding costly data structures and tailoring the move proposal distribution to the current state via a softmax over cost differences, the algorithm strikes a balance between exploration and exploitation while remaining memory efficient. Our theoretical discussion underscores the necessity of using acceptance probabilities that are consistent with the Metropolis–Hastings framework, and our numerical experiments show that this implementation reliably finds the global minimum across a range of grid sizes.

The topological analysis of the landscape revealed that the pseudorandom function exhibits a small number of deep, isolated basins (corresponding to the global and major local minima) surrounded by higher plateaus. By studying the number of connected components of sublevel and superlevel sets as a function of the

threshold, we uncovered two critical levels at which new components appear and merge, reflecting the dominant minima of the landscape. These insights support the choice of neighborhood structure and cooling schedule in our algorithm.

Future work could investigate adaptive schedules for the inverse temperature β_t based on on-line estimates of the landscape's topological complexity and extend the analysis to higher-dimensional grids or continuous domains.

References

Dimitris Bertsimas and John Tsitsiklis. "Simulated annealing." *Statistical Science* 8(1):10–15, 1993.

Herbert Edelsbrunner and John L. Harer. *Computational Topology: An Introduction*. American Mathematical Society, 2010.

Bruce Hajek. "Cooling schedules for optimal annealing." *Mathematics of Operations Research* 13(2):311–329, 1988.

Vasilios I. Manousiouthakis and Michael W. Deem. "Strict detailed balance is unnecessary in Monte Carlo simulation." *Journal of Chemical Physics* 110(6):2753–2756, 1999.

Giorgio Micaletto. *Simulated-annealing grid and sampler code*. GitHub repository. URL: https://github.com/GiorgioMB/UniversityProjects/tree/main/Course%20Related%20Projects/Computer%20Programming.