

From simulated annealing to stochastic continuation: a new trend in combinatorial optimization

Marc C. Robini · Pierre-Jean Reissman

Received: 13 May 2011 / Accepted: 24 January 2012 / Published online: 7 February 2012
© Springer Science+Business Media, LLC. 2012

Abstract Simulated annealing (SA) is a generic optimization method that is quite popular because of its ease of implementation and its global convergence properties. However, SA is widely reported to converge very slowly, and it is common practice to allow extra freedom in its design at the expense of losing global convergence guarantees. A natural way to increase the flexibility of SA is to allow the objective function and the communication mechanism to be temperature-dependent, the idea being to gradually reveal the complexity of the optimization problem and to increase the mixing rate at low temperatures. We call this general class of annealing processes stochastic continuation (SC). In the first part of this paper, we introduce SC starting from SA, and we derive simple sufficient conditions for the global convergence of SC. Our main result is interesting in two respects: first, the conditions for global convergence are surprisingly weak—in particular, they do not involve the variations of the objective function with temperature—and second, exponential cooling makes it possible to be arbitrarily close to the best possible convergence speed exponent of SA. The second part is devoted to the application of SC to the problem of producing aesthetically pleasing drawings of undirected graphs. We consider the objective function defined by Kamada and Kawai (Inf Process Lett 31(1):7–15, 1989), which measures the quality of a drawing as a weighted sum of squared differences between Euclidean and graph-theoretic inter-vertex distances. Our experiments show that SC outperforms SA with optimal communication setting both in terms of minimizing the objective function and in terms of standard aesthetic criteria.

Keywords Combinatorial optimization · Simulated annealing · Stochastic continuation · Markov chains · Monte-Carlo methods · Graph drawing

M. C. Robini (✉)
CREATIS (CNRS UMR 5220; INSERM U1044), INSA-Lyon, 69621 Villeurbanne Cedex, France
e-mail: marc.robini@creatis.insa-lyon.fr

P.-J. Reissman
Sales and E-commerce Platforms Division, Amadeus SAS, 485 route du Pin Montard,
06902 Sophia Antipolis, France
e-mail: preissman@amadeus.com

1 Introduction

1.1 Background

Simulated annealing (SA) is a generic method for combinatorial optimization that is quite popular because of its ease of implementation and its global convergence properties. It has been successfully applied to many difficult problems such as the travelling salesman [1, 36], graph partitioning [32], graph coloring [33], clique partitioning [17], graph embedding [16], task scheduling [11], image reconstruction [49, 51], image segmentation [57], and biclustering [7].

The key feature of SA is to allow uphill moves (that is, moves that increase the value of the objective function) in order to escape local minima. By analogy with the physical process of annealing in solids, uphill moves are accepted with some probability controlled by a temperature parameter that decreases monotonically to zero. As the temperature goes to zero, the invariant measure of the underlying Markov chain model concentrates on the ground states (that is, the global minima) of the objective function, and we can expect that the process converges to a ground state if the cooling is sufficiently slow. Early results [13, 25, 28] show that this is indeed the case if the temperature is inversely proportional to the logarithm of the iteration index, but this theoretical advantage is outweighed by well-known practical disadvantages: SA converges very slowly and the convergence assumptions severely limit design freedom.

Good SA algorithm design means careful selection of the cooling schedule—the most successful applications of SA use exponential cooling, which is theoretically justified in [9]—and clever construction of the candidate-solution generation mechanism (we call it the communication mechanism for short). However, many implementations of SA found in the literature use unadapted cooling schedules and crude communication mechanisms, which translates to convergence to poor local minima and sensitivity to initialization; it is therefore not surprising that SA is often abandoned in favor of other (mainly deterministic) optimization methods. The truth is that carefully designed annealing algorithms produce very good results for a wide class of problems. Yet, despite various theoretical and practical improvements over the last two decades, SA is generally much slower than deterministic methods and efficient acceleration techniques are welcome. As a result, it is common practice to relax some of the convergence conditions of SA as well as to allow extra freedom in its design, but variations on the theme of SA usually come without optimal convergence guarantees.

1.2 Contributions of the paper

1.2.1 Stochastic continuation

A natural generalization of SA is to allow the objective function and the communication mechanism to be temperature-dependent; we call the class of such algorithms stochastic continuation (SC). The first idea is to ease the annealing process by gradually revealing the complexity of the optimization problem. The second idea is to facilitate the exploration of the state space by adapting the communication mechanism to the temperature regime.

SC includes SA with temperature-dependent energy, which is studied in [22, 39, 48], and it belongs to the general class of Markov processes investigated in [18]. Yet the convergence conditions obtained in these papers strongly limit the freedom in parameterizing the objective function with temperature, and [18, 22, 39] impose impractical logarithmic cooling schedules. We show here that these limitations can be overcome while allowing the

communication mechanism to vary with temperature. Our starting point is the observation that the transitions of SC obey a large deviation principle with speed inversely proportional to the temperature, which suggests to appeal to generalized SA (GSA) theory [12,56]. This theory makes the finite state space assumption, and so do we. This is not a problem in most practical situations, as it is generally possible to increase the number of states to achieve the desired level of accuracy. Besides, the existing results on Markov Chain Monte-Carlo optimization on general state spaces—especially continuous domains—are not flexible enough to study the behavior of SC algorithms: they do not allow the energy to be time-dependent, they either require logarithmic cooling [24,27,37,38] or impose too restrictive conditions on the communication mechanism [3,59], and they do not provide relevant information on the convergence speed.

In this paper, we refine some of our results obtained in [50] by placing more emphasis on the line of thought from SA to SC and by delving more deeply into the GSA interpretation of SC—we also provide algorithmic details as well as fast methods for selecting the initial and final temperatures of the cooling schedule. We obtain a bound on the probability to exceed some threshold above the minimum objective value, and our main result states that SC with suitably adjusted exponential cooling can have a convergence speed exponent arbitrarily close to the optimal exponent of SA. Moreover, the associated conditions are surprisingly weak; in particular, they do not involve the variations of the objective function with temperature. (The reader interested in a brief overview of SC may consult [52]. Compared to this short communication, the present paper offers a comprehensive background, a better understanding of SC and its convergence properties, full proofs, practical implementation considerations, and a successful application to the graph layout problem introduced below.)

1.2.2 Graph drawing

The second contribution of this paper is the application of SC to the problem of producing aesthetically pleasing drawings of undirected, connected graphs (we refer to [35,54] for a comprehensive introduction to this subject). We confine ourselves to straight-line representations, in which each edge is mapped to a line segment in the plane. In this case, the problem reduces to that of finding a suitable mapping from the vertex set \mathcal{V} to \mathbb{R}^2 , or from \mathcal{V} to $\{1, \dots, L\}^2$ for a given integer $L \gg |\mathcal{V}|$.

There are many criteria for judging the aesthetic quality of a graph layout [46,58], but the most popular approaches are based on simple, physically-inspired models that are expressed in terms of forces acting on the objects (the vertices) of the system (the graph) or in terms of a potential energy reflecting the internal stress of the system. The layout is the result of the simulation of the relaxation of the system, which amounts to minimize a cost function defined either implicitly [21,23] or explicitly [16,34], depending on which representation is used. We consider the energy model introduced by Kamada and Kawai [34], which measures the quality of a layout as a weighted sum of squared differences between Euclidean and graph-theoretic inter-vertex distances. There are two main reasons for this choice. First, the Kamada–Kawai energy favors smooth drawings with small maximum-to-minimum edge-length ratio, small edge-length dispersion, and high amount of symmetry [6], which are three consensus aesthetic criteria. Second, although seemingly simple compared to cost functions that involve multiple constraints and multiple free parameters (such as in [16]), the Kamada–Kawai energy defines a challenging optimization problem and is therefore appropriate for comparing the performance of SA and SC. We construct a family of communication mechanisms and a family of objective functions indexed by temperature to speed up the annealing process for that particular objective function. Our experimental results show that the resulting

SC algorithm substantially outperforms both SA and the well-known Kamada–Kawai algorithm [34] in terms of objective function value, maximum-to-minimum edge-length ratio, and number of edge-crossings.

1.3 Outline

This paper is organized as follows. In Sect. 2, we review the elements of SA theory that are needed to introduce SC and to assess its theoretical performance (we refer to the survey of Nikolaev and Jacobson [41] for a comprehensive treatment of SA). The formalism of SC is presented in Sect. 3, where we study its finite-time behavior by appealing to GSA theory, and where we discuss the practical tuning of the cooling schedule. Section 4 is devoted to the application of SC to the graph drawing optimization problem discussed above. Concluding remarks are given in Sect. 5.

2 Simulated annealing

We consider the problem of finding a global minimum of an arbitrary real-valued function U defined on a finite state space Ω ; we call this objective function the energy function. We denote the ground state energy $\inf_{x \in \Omega} U(x)$ by U_{inf} , and we let $\Omega_0(U)$ be the set of global minima of U , that is, $\Omega_0(U) = \{x \in \Omega \mid U(x) = U_{\text{inf}}\}$.

2.1 Fundamentals

SA operates on an energy landscape (Ω, U, θ) defined by a symmetric and irreducible Markov matrix θ on Ω , called the communication matrix, which gives the probabilities of the possible moves for generating a candidate solution from the current solution. More precisely, $\sum_{z \in \Omega} \theta(x, z) = 1$ for all $x \in \Omega$ (that is, θ is a Markov matrix), $\theta(x, y) = \theta(y, x)$ for all $(x, y) \in \Omega^2$ (that is, θ is symmetric), and the digraph

$$(\Omega, \Delta(\theta)), \quad \Delta(\theta) = \{(x, y) \in \Omega^2 \mid y \neq x \text{ and } \theta(x, y) > 0\}, \tag{1}$$

is strongly connected (that is, θ is irreducible). In simple terms, $\Delta(\theta)$ is the set of allowed moves and the irreducibility means that any state can be reached from any other state. The most simple communication mechanisms are defined by selecting a neighborhood system \mathcal{N} on Ω (that is, $\mathcal{N} = \{\mathcal{N}(x); x \in \Omega\}$ is a collection of subsets of Ω such that (i) $x \notin \mathcal{N}(x)$, and (ii) $y \in \mathcal{N}(x) \iff x \in \mathcal{N}(y)$) and setting

$$\theta(x, y) = \begin{cases} c & \text{if } y \in \mathcal{N}(x) \\ 1 - c |\mathcal{N}(x)| & \text{if } y = x \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

where c is a positive constant such that $c \leq 1/(\sup_{x \in \Omega} |\mathcal{N}(x)|)$. More sophisticated mechanisms are constructed by weighting the allowed moves via a positive function Γ defined on $\{\{x, y\} \subset \Omega \mid y \in \mathcal{N}(x)\}$; they are of the form

$$\theta(x, y) = \begin{cases} c \Gamma(\{x, y\}) & \text{if } y \in \mathcal{N}(x) \\ 1 - c \sum_{z \in \mathcal{N}(x)} \Gamma(\{x, z\}) & \text{if } y = x \\ 0 & \text{otherwise,} \end{cases} \tag{3}$$

where $0 \leq c \leq 1/(\sup_{x \in \Omega} \sum_{z \in \mathcal{N}(x)} \Gamma(\{x, z\}))$. Note that the communication matrices (2) and (3) are clearly symmetric, and that they are irreducible if and only if the adjacency graph $(\Omega, \{\{x, y\} \subset \Omega \mid y \in \mathcal{N}(x)\})$ is connected (in both cases, $(x, y) \in \Delta(\theta)$ if and only if $y \in \mathcal{N}(x)$).

An SA process on an energy landscape (Ω, U, θ) is defined by a family $(P_\beta)_{\beta \in \mathbb{R}_+}$ of Markov matrices on Ω of the form

$$P_\beta(x, y) = \begin{cases} \theta(x, y) \mathcal{A}_\beta(x, y) & \text{if } y \neq x \\ 1 - \sum_{z \in \Omega \setminus \{x\}} P_\beta(x, z) & \text{if } y = x, \end{cases} \tag{4}$$

where the so-called acceptance probability function $\mathcal{A}_\beta : \Omega^2 \rightarrow [0, 1]$ is defined by

$$\mathcal{A}_\beta(x, y) = \exp(-\beta(U(y) - U(x))^+) \tag{5}$$

with $a^+ := \sup\{a, 0\}$. The parameter β plays the role of an inverse temperature, and $\mathcal{A}_\beta(x, y)$ is the probability to accept the move from the current solution x to the candidate solution y at temperature β^{-1} . Other acceptance probability functions are possible (we then speak of an hill-climbing process [31]), but it is shown in [53] that (5) is the unique form such that (i) $\mathcal{A}_\beta(x, y) = 1$ if $U(y) \leq U(x)$, (ii) \mathcal{A}_β depends uniformly on the energy difference between the current and candidate solutions, and (iii) the Markov chain $(X_n)_{n \in \mathbb{N}}$ with transitions $\mathbb{P}(X_n = y \mid X_{n-1} = x) = P_\beta(x, y)$ is reversible.

We call a positive real sequence $(\beta_n)_{n \in \mathbb{N}^*}$ a cooling sequence if it is non-decreasing and if $\lim_{n \rightarrow +\infty} \beta_n = +\infty$. Given such a sequence, an SA algorithm on (Ω, U, θ) is a discrete-time, non-homogeneous Markov chain $(X_n)_{n \in \mathbb{N}}$ with transitions $\mathbb{P}(X_n = y \mid X_{n-1} = x) = P_{\beta_n}(x, y)$. We use the notation $\text{SA}(\Omega, U, \theta, (\beta_n))$ for short. In practice, a finite-time realization $(x_n)_{n \in \llbracket 0, N \rrbracket}$ ($\llbracket 0, N \rrbracket$ is a shorthand notation for $\{0, \dots, N\}$) of an annealing chain $\text{SA}(\Omega, U, \theta, (\beta_n))$ is generated as follows:

```

pick an initial state  $x_0 \in \Omega$ ;
for  $n = 1$  to  $N$  do
    draw a state  $y$  from the probability distribution  $\theta(x_{n-1}, \cdot)$  on  $\Omega$ ;
    set  $x_n \leftarrow x_{n-1}$ ;
    set  $\delta \leftarrow U(y) - U(x_{n-1})$ ;
    if  $\delta \leq 0$  then set  $x_n \leftarrow y$ ;
    else set  $x_n \leftarrow y$  with probability  $\exp(-\beta_n \delta)$ ;
    end(if)
end(for)
    
```

The Markov matrix P_β inherits the irreducibility of θ for any β ; thus, since Ω is finite, P_β has a unique and positive invariant measure, which we denote by μ_β . From the symmetry of θ , we have $\exp(-\beta U(x))P_\beta(x, y) = \exp(-\beta U(y))P_\beta(y, x)$ for all (x, y) , and therefore

$$\mu_\beta(x) \propto \exp(-\beta U(x)). \tag{6}$$

In other words, μ_β is the Gibbs distribution with energy U at temperature β^{-1} . When β increases to infinity, μ_β concentrates around the ground states and tends to the uniform distribution on $\Omega_0(U)$, that is,

$$\lim_{\beta \rightarrow +\infty} \mu_\beta(x) = \begin{cases} 1/|\Omega_0(U)| & \text{if } x \in \Omega_0(U) \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

At this point, it is natural to question the need for cooling. Indeed, we can think of searching for the global minima by Metropolis sampling, which consists in simulating an homogeneous Markov chain with transitions matrix P_β for a fixed β and keeping the lowest energy state found during the simulation. Metropolis sampling has interesting finite-time convergence properties [10, 19, 44, 45], and some experimental results show that it can perform comparably to SA if the temperature is chosen correctly [14, 20]. Unfortunately, there is no general approach to choosing a fixed temperature value appropriate to a given optimization problem. If we want to be reasonably sure of finding a good solution, we have to choose β large enough so that μ_β is sharply peaked around the ground states. On the other hand, the larger β , the less mobile the Metropolis chain, and hence the more likely it is to get stuck in poor local minima. The truth is that it is generally not possible to design an efficient communication mechanism for rapid mixing at low temperatures. This motivates the idea of annealing: if the cooling sequence $(\beta_n)_n$ increases sufficiently slowly, then we can expect that the law of X_n is close to μ_{β_n} so as to achieve asymptotic optimality, that is,

$$\lim_{n \rightarrow +\infty} \inf_{x \in \Omega} \mathbf{P}(X_n \in \Omega_0(U) \mid X_0 = x) = 1. \tag{8}$$

From this perspective, SA can be viewed as an acceleration technique for Metropolis sampling.

2.2 Main convergence results

The most well-known asymptotic convergence result for SA is due to Hajek [28], who showed that (8) holds if and only if

$$\sum_{n=1}^{+\infty} \exp(-\beta_n H_c) = +\infty, \tag{9}$$

where H_c is the maximum energy barrier separating a non-optimal state from a ground state. The constant H_c is called the critical depth of the energy landscape. Formally,

$$H_c = \sup_{x \in \Omega \setminus \Omega_0(U)} H(x), \tag{10}$$

where $H(x)$ —the depth of x —is defined as follows:

$$H(x) = \inf_{y \in \Omega_0(U)} h(x, y) - U(x) \tag{11}$$

$$\text{with } h(x, y) = \inf_{(x_i)_{i=1}^m \in \Pi_\theta(x, y)} \sup_{i \in \llbracket 1, m \rrbracket} U(x_i), \tag{12}$$

where $\Pi_\theta(x, y)$ denotes the set of paths from x to y in $(\Omega, \Delta(\theta))$.

Hajek’s result readily implies that logarithmic cooling sequences of the form $\beta_n = \beta_0 \ln n$ are asymptotically optimal if $0 < \beta_0 \leq 1/H_c$. Notable refinements were given by Chiang and Chow [13] and by Catoni [8], who provided necessary and sufficient conditions for the limit distribution of the annealing chain to give a strictly positive mass to any $x \in \Omega_0(U)$ and for strong ergodicity, but these results do not provide information about the status of SA after a finite number of steps. The finite-time behavior of SA with logarithmic cooling was investigated by Desai and Rao [19], who derived an upper bound on the number $N(\eta, p)$ of iterations needed to ensure that a state with energy lower than $U_{\text{inf}} + \eta$ has been encountered with probability p . However, in practice, logarithmic cooling yields extremely slow convergence and is usually replaced by exponential cooling, the justification of which is outlined below.

Let $(\beta_n^N)_{n \in \llbracket 1, N \rrbracket}$ be a finite cooling sequence, and consider the convergence measure $\mathcal{M}(N)$ of the finite-time algorithm $(X_n^N)_{n \in \llbracket 0, N \rrbracket} = \text{SA}(\Omega, U, \theta, (\beta_n^N))$ defined by

$$\mathcal{M}(N) = \sup_{x \in \Omega} \mathbb{P}(X_N^N \notin \Omega_0(U) \mid X_0^N = x). \tag{13}$$

It is shown in [9] that as the horizon N increases, $\mathcal{M}(N)$ cannot decrease faster than some optimal exponent of N^{-1} . More precisely, letting $\mathcal{B}(N)$ be the set of finite cooling sequences of length N (that is, $\mathcal{B}(N) = \{(\beta_n^N) \mid 0 \leq \beta_1^N \leq \dots \leq \beta_N^N\}$), we have

$$\lim_{N \rightarrow +\infty} \sup_{(\beta_n^N) \in \mathcal{B}(N)} - \frac{\ln \mathcal{M}(N)}{\ln N} \leq \frac{1}{D}, \tag{14}$$

where the constant D , called the difficulty of the energy landscape, is the maximum ratio of the depth to the energy level above the ground state energy:

$$D = \sup_{x \in \Omega \setminus \Omega_0(U)} \frac{H(x)}{U(x) - U_{\text{inf}}}. \tag{15}$$

Furthermore, the upper bound $1/D$ in (14) is sharp, as $\ln \mathcal{M}(N) \sim \ln(N^{-1/D})$ for some families $\{(\beta_n^N); N \in \mathbb{N}^*\}$ of cooling sequences of the form

$$\beta_n^N = \beta_0 \exp(nf(N)) \quad \text{with} \quad f(N) \sim N^{-1} \ln N, \tag{16}$$

where β_0 is independent of the horizon N . This rigorous justification for exponential cooling is a direct consequence of Theorem 8.1 in [9], where it is also established that there exists piecewise logarithmic sequences such that $\mathcal{M}(N) \leq CN^{-1/D}$ for some positive constant C (however, these sequences depend strongly on the hierarchical structure of the energy landscape and their identification is intractable for problems of practical size). On the experimental side, the optimal cooling sequence attached to a particular optimization problem may be neither logarithmic nor exponential [14], but exponential cooling remains particularly attractive in practice because, contrary to other cooling strategies, it is uniformly robust with respect to the energy landscape.

It can be checked that the supremum in the definition (15) of the difficulty of the energy landscape can be taken over the set $\Omega_{\text{loc}}(U, \theta) \setminus \Omega_0(U)$, where $\Omega_{\text{loc}}(U, \theta)$ denotes the set of local minima of the energy landscape (Ω, U, θ) :

$$\Omega_{\text{loc}}(U, \theta) = \{x \in \Omega \mid \forall y \in \Omega, \theta(x, y) > 0 \implies U(x) \leq U(y)\}. \tag{17}$$

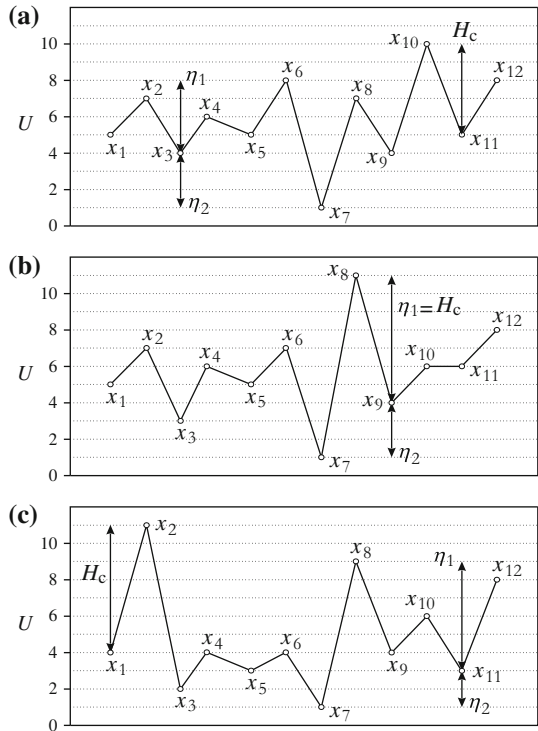
Therefore, the above finite-time convergence properties are consistent with the intuitive understanding of annealing: SA performs poorly if the energy landscape has low-energy non-global minima and if these minima are separated from the ground states by high energy barriers. By way of illustration, Fig. 1 shows three simple energy landscapes with increasing difficulty. In each case, $\Omega = \{x_i; i \in \llbracket 1, 12 \rrbracket\}$, $U(\Omega) \subset \mathbb{N}$, and $\theta(x, y) > 0$ if and only if $(x, y) = (x_i, x_{i+1})$ or (x_i, x_{i-1}) . The quantities η_1 and η_2 are defined by

$$\eta_1 = H(x^*) \quad \text{and} \quad \eta_2 = U(x^*) - U_{\text{inf}} \tag{18}$$

$$\text{with } x^* \in \arg \sup_{x \in \Omega \setminus \Omega_0(U)} \frac{H(x)}{U(x) - U_{\text{inf}}}, \tag{19}$$

and thus $D = \eta_1/\eta_2$. As exemplified by Fig. 1a and c, the non-global minimum with maximum depth does not necessarily coincide with the argument of the supremum in the definition of the difficulty. In fact, the ordering of the non-global minima in terms of the depth H is

Fig. 1 Energy landscapes with increasing difficulty $D = \eta_1/\eta_2$: **a** $D = \frac{4}{3}$; **b** $D = \frac{7}{2}$; **c** $D = 3$



generally not the same as that defined by $H/(U - U_{\text{inf}})$, and thus the notion of a local basin of attraction differs between asymptotic and finite-time convergence theories.

We end this section by noting that the finite-time convergence theory also sheds new light on the benefits of SA over Metropolis sampling. From [10], the optimal convergence speed exponent of the Metropolis algorithm is $1/D_M$ with

$$D_M = \frac{H_c}{\inf_{x \in \Omega \setminus \Omega_0(U)} U(x) - U_{\text{inf}}}. \tag{20}$$

Letting $\Omega_{\text{loc}}^\dagger := \Omega_{\text{loc}}(U, \theta) \setminus \Omega_0(U)$ be the set of non-global minima of (Ω, U, θ) , we have $D < D_M$ if and only if

$$\left(\exists x \in \Omega, U_{\text{inf}} < U(x) < \inf_{y \in \Omega_{\text{loc}}^\dagger} U(y) \right) \\ \vee \left(\forall x \in \Omega_{\text{loc}}^\dagger, H(x) = \sup_{y \in \Omega_{\text{loc}}^\dagger} H(y) \implies U(x) > \inf_{y \in \Omega_{\text{loc}}^\dagger} U(y) \right).$$

In other words, SA is potentially faster than Metropolis sampling if there is a state $x \notin \Omega_0(U)$ with smaller energy than any non-global minimum or if the set of non-global minima with maximum depth is disjoint from the set of non-global minima with minimum energy. For example, going back to Fig. 1, the Metropolis difficulty is greater than D in all three cases: (a) $D_M = \frac{5}{3}$, (b) $D_M = \frac{7}{2}$, and (c) $D_M = 7$ (the corresponding values of the acceleration exponent $1/D - 1/D_M$ are $\frac{3}{20}$, $\frac{1}{7}$ and $\frac{4}{21}$, respectively).

3 Stochastic continuation

The relationship between the convergence rate of SA and the difficulty of the energy landscape suggests a possible acceleration by making the energy temperature-dependent. The idea is to guide the hierarchical search performed by SA—and thus to reduce the risk of getting stuck in undesirable basins of attraction—by replacing the energy function U with a sequence $(U_n)_n$ of functions converging pointwise to U and such that the difficulty of (Ω, U_n, θ) increases with n . In a similar vein, since static communication is generally efficient over only a small range of temperatures, another potential improvement is to adapt the communication mechanism to the temperature regime. This leads to an important class of generalized SA (GSA) algorithms in which the temperature controls not only the acceptance rate of uphill moves, but also the energy function and the communication matrix. We call it SC by analogy with deterministic continuation methods, in which the minima are tracked by computing successive approximate solutions from a parameterized energy that tends to the objective function as the iterations increase (see, for instance, [4, 40, 42]).

3.1 Definition and basic idea

We define an SC process with target energy landscape (Ω, U, q) to be a family $(Q_\beta)_{\beta \in \mathbb{R}_+}$ of Markov matrices on Ω of the form

$$Q_\beta(x, y) = \begin{cases} q_\beta(x, y) \exp(-\beta(U_\beta(y) - U_\beta(x))^+) & \text{if } y \neq x \\ 1 - \sum_{z \in \Omega \setminus \{x\}} Q_\beta(x, z) & \text{if } y = x, \end{cases} \tag{21}$$

$$\text{with } \lim_{\beta \rightarrow +\infty} U_\beta(x) = U(x) \quad \text{and} \quad \lim_{\beta \rightarrow +\infty} q_\beta(x, y) = q(x, y).$$

Given such a family together with a cooling sequence $(\beta_n)_{n \in \mathbb{N}^*}$, we call a Markov chain $(X_n)_{n \in \mathbb{N}}$ on Ω with transitions $\mathbb{P}(X_n = y \mid X_{n-1} = x) = Q_{\beta_n}(x, y)$ an SC algorithm, and we denote it by $\text{SC}(\Omega, (U_\beta), (q_\beta), (\beta_n))$ —we use the notation $(\Omega, (Q_\beta))$ or $(\Omega, (U_\beta), (q_\beta))$ for the underlying SC process. The family of functions $(U_\beta : \Omega \rightarrow \mathbb{R})_\beta$ is called the continuation scheme, and the family of Markov matrices $(q_\beta : \Omega^2 \rightarrow [0, 1])_\beta$ is called the communication scheme.

We assume that the limit communication matrix q is irreducible, as there is otherwise no guarantee to reach a ground state of the target energy U . The basic idea of SC is easy to explain under the additional assumption that q_β is symmetric for β sufficiently large (this restriction will be relaxed in the next section). Indeed, Proposition 1 below states that in this case the invariant measure ν_β of Q_β is a Gibbs distribution which concentrates on the set of global minima of U as $\beta \rightarrow +\infty$. Consequently, similarly to SA, if the cooling sequence does not increase too fast, the law of X_n should stay close enough to ν_{β_n} to expect convergence to an optimum. This gives the go-ahead for studying the global convergence properties of SC.

Proposition 1 *Let $(\Omega, (Q_\beta))$ be an SC process with q irreducible and q_β symmetric for β sufficiently large. Then, there exists $\beta^* \geq 0$ such that for any $\beta \geq \beta^*$, Q_β is irreducible and its unique invariant measure ν_β satisfies*

$$\nu_\beta(x) \propto \exp(-\beta U_\beta(x)) \tag{22}$$

$$\text{and } \lim_{\beta \rightarrow +\infty} \nu_\beta(\Omega_0(U)) = 1. \tag{23}$$

Proof Since Ω is finite, there exists $\beta^* \geq 0$ such that for any $\beta \geq \beta^*$, q_β is symmetric and $\Delta(q) \subseteq \Delta(q_\beta)$, where Δ is defined in (1).

Let $\beta \geq \beta^*$. The Markov matrix q_β inherits the irreducibility of q , and thus, since $\Delta(q_\beta) \subseteq \Delta(Q_\beta)$, Q_β is irreducible. It follows that Q_β has a unique invariant measure. Let ν_β be the probability distribution on Ω defined by

$$\nu_\beta(x) = \frac{\exp(-\beta U_\beta(x))}{Z_\beta(\Omega)}, \quad Z_\beta(\mathcal{A}) := \sum_{z \in \mathcal{A}} \exp(-\beta U_\beta(z)).$$

For any $(x, y) \in \Omega^2$ such that $x \neq y$, we have

$$\nu_\beta(x)Q_\beta(x, y) = q_\beta(x, y) \frac{\exp(-\beta \sup\{U_\beta(x), U_\beta(y)\})}{Z_\beta(\Omega)}$$

and thus

$$\nu_\beta(x)Q_\beta(x, y) = \nu_\beta(y)Q_\beta(y, x)$$

by the symmetry of q_β . In other words, Q_β is reversible with respect to ν_β , which implies that ν_β is invariant for Q_β .

Let us put $\Omega_0 := \Omega_0(U)$ and $\overline{\Omega_0} := \Omega \setminus \Omega_0(U)$, and let f_0 and f_1 be the real-valued functions on \mathbb{R}_+ defined by

$$f_0(\beta) = \sup_{z \in \Omega_0} U_\beta(z) \quad \text{and} \quad f_1(\beta) = \inf_{z \in \overline{\Omega_0}} U_\beta(z).$$

We have

$$Z_\beta(\Omega_0) \geq |\Omega_0| \exp(-\beta f_0(\beta)), \quad Z_\beta(\overline{\Omega_0}) \leq |\overline{\Omega_0}| \exp(-\beta f_1(\beta)),$$

and $\lim_{\beta \rightarrow +\infty} (f_1 - f_0)(\beta) = \inf_{z \in \overline{\Omega_0}} U(z) - U_{\inf} > 0$.

Therefore,

$$\frac{Z_\beta(\overline{\Omega_0})}{Z_\beta(\Omega_0)} \in O_{\beta \rightarrow +\infty}(\exp(-\beta(f_1 - f_0)(\beta))) \xrightarrow{\beta \rightarrow +\infty} 0$$

and hence

$$\nu_\beta(x) \underset{\beta \rightarrow +\infty}{\sim} \frac{\exp(-\beta U_\beta(x))}{Z_\beta(\Omega_0)} \in O_{\beta \rightarrow +\infty}(\exp(-\beta(U_\beta(x) - f_0(\beta)))). \tag{24}$$

Now, for any $x \in \overline{\Omega_0}$,

$$\lim_{\beta \rightarrow +\infty} (U_\beta(x) - f_0(\beta)) = U(x) - U_{\inf} > 0,$$

and it follows from (24) that $\lim_{\beta \rightarrow +\infty} \nu_\beta(\overline{\Omega_0}) = 0$. □

Remark 1 SC is seemingly related to generalized hill-climbing (GHC) [31], which includes SA as a special case. The acceptance probability function of a GHC process is of the form

$$\mathcal{A}_\beta(x, y) = \mathbf{P}(S_\beta(x, y) \geq U(y) - U(x)), \tag{25}$$

where $S_\beta(x, y)$ is a non-negative random variable whose density function is parameterized by β and depends on (x, y) (in the case of SA, $S_\beta(x, y) = -\beta^{-1} \ln \kappa_x$, where κ_x is a random variable uniformly distributed on $(0, 1)$). Strictly speaking, SA with temperature-dependent

energy—and hence SC—is not included in GHC. Indeed, if there is a pair $(x, y) \in \Omega^2$ such that $U(y) = U(x)$ and $U_\beta(y) > U_\beta(x)$ for all β , then for such a pair, the acceptance probability of GHC is $P(S_\beta(x, y) \geq 0) = 1$ whereas that of SC is $\exp(-\beta(U_\beta(y) - U_\beta(x))) < 1$. We refer the reader interested in GHC to the work of Johnson and Jacobson [31] and Jacobson and Yucsan [30]. Both papers examine the asymptotic convergence of GHC: [31] discusses the standard homogeneous Markov chain approach, and [30] provides necessary and sufficient conditions that generalize the result of Hajek discussed in Sect. 2.2.

Remark 2 SC includes the class of so-called compressed SA (CSA) algorithms, in which a variable penalty multiplier approach is used to solve constrained optimization problems [43,47]. The objective of a CSA algorithm is to find a global minimum of an energy $U : \Omega \rightarrow \mathbb{R}$ over a set $\Omega_g \subset \Omega$ of the form $\{x \in \Omega \mid g(x) = \inf_{y \in \Omega} g(y)\}$, where $g : \Omega \rightarrow \mathbb{R}$ is non-constant. CSA is similar to SA except for the acceptance probability function which is controlled not only by temperature, but also by a positive parameter λ acting as a Lagrange multiplier:

$$A_{\beta,\lambda}(x, y) = \exp\left(-\beta((U(y) + \lambda g(y)) - (U(x) + \lambda g(x)))^+\right). \tag{26}$$

Simply speaking, U is replaced by the augmented energy $U + \lambda g$. The rationale behind this is that, since Ω is finite, the set of global minima of $U + \lambda g$ on Ω and the set of global minima of U on Ω_g are equal if λ is sufficiently large. The idea of CSA is to make λ increase smoothly with increasing β so as to guarantee convergence to a feasible solution while moving freely through the whole state space at high temperatures. Writing λ as a function Λ of β , we see that CSA belongs to the class of SC algorithms with fixed communication and with continuation scheme of the form $U_\beta = U + \Lambda(\beta)g$. Therefore, the finite-time convergence properties of SC derived in the next section shed new light on the behavior of CSA.

3.2 Finite-time convergence

SC is included in the general class of Markov processes investigated in [18]; it is an extension of SA with temperature-dependent energy, the behavior of which is studied in [22,39] for the asymptotic case and in [48] for the finite-time case. The convergence results in [18,48] require that

$$\sup_{(x,\beta) \in \Omega \times \mathbb{R}_+} \beta |U_\beta(x) - U(x)| < +\infty, \tag{27}$$

while it is assumed in [22,39] that there exists $a > 0$ such that

$$\sup_{(x,n) \in \Omega \times \mathbb{N}^*} n^a |U_{\beta_{n+1}}(x) - U_{\beta_n}(x)| < +\infty. \tag{28}$$

Both conditions (27) and (28) significantly limit the freedom in parameterizing the energy with temperature. Furthermore, the convergence results in [18,22,39] are limited to the asymptotic case and involve impractical logarithmic cooling sequences. In this section, we show that these limitations can be overcome while allowing the communication mechanism to vary with temperature. More specifically, our analysis of the finite-time convergence of SC refines our previous results obtained in [50] in three ways: first, we relax the condition that the limit communication mechanism can rest anywhere (that is, it is no longer required that $q(x, x) > 0$ for all x); second, we give an upper bound on the probability that the final energy level exceeds some given threshold above the ground state energy; and third, we provide a

more precise description of the families of cooling sequences that allow to approach the optimal convergence speed exponent of SA.

We start from the basic observation that SC and SA behave similarly at low temperatures in the sense that

$$\forall(x, y) \in \Delta(q), \quad \lim_{\beta \rightarrow +\infty} \frac{-\ln Q_\beta(x, y)}{\beta} = \lim_{\beta \rightarrow +\infty} \frac{-\ln P_\beta(x, y)}{\beta} = (U(y) - U(x))^+. \tag{29}$$

In other words, for any $(x, y) \in \Omega^2$ such that $y \neq x$ and $q(x, y) > 0$, $Q_\beta(x, y)$ satisfies a large deviation principle with speed β and rate $(U(y) - U(x))^+$, which suggests to use the GSA theory developed in [12, 15, 56].

A GSA process on Ω is defined by a family $(\Theta_\beta)_{\beta \in \mathbb{R}_+}$ of Markov matrices on Ω satisfying a large deviation assumption with speed β and with irreducible rate function $J : \Omega^2 \rightarrow \mathbb{R}_+ \cup \{+\infty\}$; that is,

$$\forall(x, y) \in \Omega^2, \quad \lim_{\beta \rightarrow +\infty} \frac{-\ln \Theta_\beta(x, y)}{\beta} = J(x, y) \tag{30}$$

(with the convention that $\ln 0 = -\infty$) and the digraph

$$(\Omega, \tilde{\Delta}(J)), \quad \tilde{\Delta}(J) = \{(x, y) \in \Omega^2 \mid y \neq x \text{ and } J(x, y) < +\infty\} \tag{31}$$

is strongly connected. Given a GSA process $(\Omega, (\Theta_\beta))$ and a cooling sequence $(\beta_n)_n$, we call a Markov chain $(X_n)_{n \in \mathbb{N}}$ on Ω with transitions $\mathbb{P}(X_n = y \mid X_{n-1} = x) = \Theta_{\beta_n}(x, y)$ a GSA algorithm; we denote it by $\text{GSA}(\Omega, (\Theta_\beta), (\beta_n))$.

Clearly, if q is irreducible, then any rate function J such that $J(x, y) = (U(y) - U(x))^+$ for all $(x, y) \in \Delta(q)$ is irreducible. Consequently, from (29), we can appeal to GSA theory provided we can find conditions so that for any $(x, y) \notin \Delta(q)$, $-\beta^{-1} \ln Q_\beta(x, y)$ has a limit in $\mathbb{R}_+ \cup \{+\infty\}$ as $\beta \rightarrow +\infty$. The following proposition makes this precise.

Proposition 2 *Let $(\Omega, (Q_\beta))$ be an SC process with communication scheme $(q_\beta)_\beta$ and with target energy landscape (Ω, U, q) . Assume that*

(A1) *q is irreducible,*

$$(A2) \quad \forall(x, y) \in \Omega^2, \quad q(x, y) = 0 \implies \lim_{\beta \rightarrow +\infty} \frac{-\ln q_\beta(x, y)}{\beta} = \begin{cases} +\infty & \text{if } y \neq x \\ 0 & \text{if } y = x. \end{cases}$$

Then $(\Omega, (Q_\beta))$ is a GSA process with rate function

$$J(x, y) = \begin{cases} (U(y) - U(x))^+ & \text{if } q(x, y) > 0 \text{ or } y = x \\ +\infty & \text{otherwise.} \end{cases} \tag{32}$$

Proof The irreducibility of J readily follows from the irreducibility of q . Taking (29) into account, we just have to show that

$$\lim_{\beta \rightarrow +\infty} \frac{-\ln Q_\beta(x, y)}{\beta} = \begin{cases} +\infty & \text{if } q(x, y) = 0 \text{ and } y \neq x \\ 0 & \text{if } y = x. \end{cases}$$

If $q(x, y) = 0$ and $y \neq x$, then

$$\begin{aligned} \frac{-\ln Q_\beta(x, y)}{\beta} &= \frac{-\ln q_\beta(x, y)}{\beta} + (U_\beta(y) - U_\beta(x))^+ \\ &\xrightarrow{\beta \rightarrow +\infty} (+\infty) + (U(y) - U(x))^+ = +\infty. \end{aligned}$$

If $y = x$, then

$$1 \geq Q_\beta(x, x) \geq 1 - \sum_{z \neq x} q_\beta(x, z) = q_\beta(x, x)$$

and hence

$$0 \leq \frac{-\ln Q_\beta(x, x)}{\beta} \leq \frac{-\ln q_\beta(x, x)}{\beta} \xrightarrow{\beta \rightarrow +\infty} 0.$$

□

Let $(\Omega, (\Theta_\beta))$ be a GSA process with rate function J . Since Ω is finite and J is irreducible, it follows from (30) that for β large enough, Θ_β is irreducible and hence has a unique invariant measure ϑ_β . We know from [12] that there is a function $V : \Omega \rightarrow \mathbb{R}_+$ such that

$$\forall x \in \Omega, \quad \lim_{\beta \rightarrow +\infty} \frac{-\ln \vartheta_\beta(x)}{\beta} = V(x) - V_{\text{inf}}, \tag{33}$$

where $V_{\text{inf}} = \inf_{x \in \Omega} V(x)$; in other words, $(\vartheta_\beta)_\beta$ satisfies a large deviation principle with speed β and with rate function $V - V_{\text{inf}}$. The function V is called the virtual energy and is defined by

$$\forall x \in \Omega, \quad V(x) = \inf_{T \in \mathcal{T}(x)} \sum_{(z,t) \in \mathcal{E}_T} J(z, t), \tag{34}$$

where $\mathcal{T}(x)$ is the set of directed trees $T = (\Omega, \mathcal{E}_T)$, $\mathcal{E}_T \subset \Omega^2$, with root x and whose edges are directed towards x (that is, $d_T^+(y) = 1$ for all $y \neq x$ and $d_T^+(x) = 0$, where $d_T^+(z)$ is the outdegree of vertex z in T). Similarly to standard SA theory, the triplet (Ω, V, J) defines an energy landscape—call it the virtual energy landscape—and is accompanied with a critical depth \tilde{H}_c and a difficulty \tilde{D} ; these two constants are defined by

$$\tilde{H}_c = \sup_{x \in \Omega \setminus \Omega_0(V)} \tilde{H}(x) \tag{35}$$

$$\text{and } \tilde{D} = \sup_{x \in \Omega \setminus \Omega_0(V)} \frac{\tilde{H}(x)}{V(x) - V_{\text{inf}}}, \tag{36}$$

where $\tilde{H}(x)$ denotes the depth of x in (Ω, V, J) :

$$\tilde{H}(x) = \inf_{y \in \Omega_0(V)} \tilde{h}(x, y) - V(x) \tag{37}$$

$$\text{with } \tilde{h}(x, y) = \inf_{(x_i)_{i=1}^m \in \tilde{\Pi}_J(x, y)} \sup_{i \in \llbracket 1, m-1 \rrbracket} (V(x_i) + J(x_i, x_{i+1})), \tag{38}$$

where $\tilde{\Pi}_J(x, y)$ is the set of paths from x to y in $(\Omega, \tilde{\Delta}(J))$.

From (33), ϑ_β concentrates on the set $\Omega_0(V)$ of global minima of the virtual energy as $\beta \rightarrow +\infty$, which is to say that V plays the role of an objective function for the GSA process $(\Omega, (\Theta_\beta))$. Hence, an SC process that fits into the GSA framework can potentially minimize its target energy U if $\Omega_0(V) \subseteq \Omega_0(U)$. Proposition 3 gives a simple condition for this to be the case.

Proposition 3 *Let $(\Omega, (Q_\beta))$ be an SC process with target energy landscape (Ω, U, q) and satisfying assumptions (A1) and (A2). If*

(A3) $\Delta(q)$ is symmetric (that is, $\forall(x, y) \in \Omega^2, q(x, y) > 0 \iff q(y, x) > 0$), then the virtual energy V of $(\Omega, (Q_\beta))$ satisfies

$$V - V_{\inf} = U - U_{\inf}. \tag{39}$$

Proof Let $(x, y) \in \Omega^2$ such that $x \neq y$. For any $T \in \mathcal{T}(x)$, there is a unique path $\pi_T(x, y)$ from y to x in T : $\pi_T(x, y)$ is a finite sequence $(x_i)_{i=1}^m$ of distinct states such that $x_1 = y, x_m = x$ and $(x_i, x_{i+1}) \in \mathcal{E}_T$ for all $i \in \llbracket 1, m - 1 \rrbracket$. Assume that $\pi_T(x, y)$ is J -admissible, that is, $J(x_i, x_{i+1}) < +\infty$ for all i . Then, for any i , we have $q(x_i, x_{i+1}) > 0$ from (32), hence $q(x_{i+1}, x_i) > 0$ by (A3), and thus

$$\begin{aligned} J(x_i, x_{i+1}) - J(x_{i+1}, x_i) &= (U(x_{i+1}) - U(x_i))^+ - (U(x_i) - U(x_{i+1}))^+ \\ &= U(x_{i+1}) - U(x_i). \end{aligned}$$

Consequently,

$$\begin{aligned} \sum_{\substack{(z,t) \in \mathcal{E}_T : \\ z \in \pi_T(x,y)}} J(z, t) &= \sum_{i=1}^{m-1} J(x_i, x_{i+1}) \\ &= \sum_{i=1}^{m-1} (U(x_{i+1}) - U(x_i)) + \sum_{i=1}^{m-1} J(x_{i+1}, x_i) \\ &= U(x) - U(y) + \sum_{\substack{(z,t) \in \mathcal{E}_T : \\ z \notin \pi_T(x,y)}} J(t, z). \end{aligned}$$

Since J is irreducible, there exists $T \in \mathcal{T}(x)$ such that

$$J(T) := \sum_{(z,t) \in \mathcal{E}_T} J(z, t) < +\infty,$$

and hence such that $\pi_T(x, y)$ is J -admissible. It follows that

$$\begin{aligned} V(x) + U(y) &= \inf_{\substack{T \in \mathcal{T}(x) : \\ J(T) < +\infty}} J(T) + U(y) \\ &= \inf_{\substack{T \in \mathcal{T}(x) : \\ J(T) < +\infty}} \left(\sum_{\substack{(z,t) \in \mathcal{E}_T : \\ z \notin \pi_T(x,y)}} J(z, t) + \sum_{\substack{(z,t) \in \mathcal{E}_T : \\ z \in \pi_T(x,y)}} J(z, t) + U(y) \right) \\ &= \inf_{\substack{T \in \mathcal{T}(x) : \\ J(T) < +\infty}} \left(\sum_{\substack{(z,t) \in \mathcal{E}_T : \\ z \notin \pi_T(x,y)}} J(z, t) + \sum_{\substack{(z,t) \in \mathcal{E}_T : \\ z \in \pi_T(x,y)}} J(t, z) + U(x) \right). \end{aligned}$$

Consider the mapping $\zeta : \mathcal{T}(x) \rightarrow \mathcal{T}(y)$ that reverses the orientation of each edge in the path from y to x , that is, $\zeta(T) = (\Omega, \mathcal{E}_\zeta(T))$ with

$$\mathcal{E}_\zeta(T) = \{(z, t) \in \mathcal{E}_T \mid z \notin \pi_T(x, y)\} \cup \{(t, z) \mid (z, t) \in \mathcal{E}_T \text{ and } z \in \pi_T(x, y)\}.$$

ζ is clearly one-to-one. Moreover, by (32) and (A3), we have $J(z, t) < +\infty \iff J(t, z) < +\infty$, and thus $J(T) < +\infty \iff J \circ \zeta(T) < +\infty$. Therefore,

$$\begin{aligned} V(x) + U(y) &= \inf_{\substack{T \in \mathcal{T}(x): \\ J(T) < +\infty}} \left(\sum_{(z,t) \in \mathcal{E}_\zeta(T)} J(z, t) + U(x) \right) \\ &= \inf_{\substack{T \in \mathcal{T}(x): \\ J \circ \zeta(T) < +\infty}} J \circ \zeta(T) + U(x) \\ &= \inf_{\substack{T \in \mathcal{T}(y): \\ J(T) < +\infty}} J(T) + U(x) \\ &= V(y) + U(x). \end{aligned}$$

In particular, given $y_0 \in \Omega_0(U)$, we have

$$V(x) - V(y_0) = U(x) - U_{\text{inf}} \geq 0$$

for all x , and hence $V(y_0) = V_{\text{inf}}$. □

From Propositions 2 and 3, and from (33), any SC process satisfying assumptions (A1)–(A3) is a GSA process whose limit invariant measure is concentrated on the set of global minima of the target energy. Therefore, the convergence properties of SC as a global optimization approach can be derived from the properties of GSA. We will use the convergence result stated in the following theorem, which gives an asymptotic bound on the probability of error of GSA for suitably adjusted piecewise-constant exponential cooling sequences.

Theorem 1 ([12]) *Let $(\Omega, (\Theta_\beta))$ be a GSA process with virtual energy landscape (Ω, V, J) . For any positive reals A, ξ and η , consider the family of finite-time algorithms*

$$\{ (X_n^{\sigma, K})_{n \in \llbracket 0, \sigma K \rrbracket} = \text{GSA}(\Omega, (\Theta_\beta), (\beta_n^{\sigma, K})); \sigma, K \in \mathbb{N}^* \}$$

with cooling sequence

$$\beta_n^{\sigma, K} = \frac{\ln K}{A} \left(\frac{\xi}{\eta} \right)^{\frac{1}{\sigma} \lfloor \frac{n-1}{K} \rfloor}, \quad n \in \llbracket 1, \sigma K \rrbracket. \tag{40}$$

If $A > \tilde{H}_c$, $\xi > \tilde{H}_c / \tilde{D}$, and $\eta < \xi$, then, for any σ ,

$$\liminf_{K \rightarrow +\infty} - \frac{\ln \sup_{x \in \Omega} \mathbf{P}(V(X_{\sigma K}^{\sigma, K}) \geq V_{\text{inf}} + \eta \mid X_0^{\sigma, K} = x)}{\ln(\sigma K)} \geq \frac{1}{\tilde{D}} \left(\frac{\eta}{\xi} \right)^{\frac{1}{\sigma}}. \tag{41}$$

Let $(\Omega, (U_\beta), (q_\beta))$ be an SC process satisfying assumptions (A1)–(A3), and let (Ω, U, q) and (Ω, V, J) be the associated target and virtual energy landscapes, respectively. From Propositions 2 and 3, we have $\Omega_0(V) = \Omega_0(U)$, $\tilde{\Delta}(J) = \tilde{\Delta}(q)$ (hence $\tilde{\Pi}_J(x, y) = \tilde{\Pi}_q(x, y)$ for all $(x, y) \in \Omega^2$ such that $x \neq y$), and for any $(z, t) \in \tilde{\Delta}(J)$,

$$\begin{aligned} V(z) + J(z, t) &= U(z) - U_{\text{inf}} + V_{\text{inf}} + (U(t) - U(z))^+ \\ &= \sup\{U(z), U(t)\} - U_{\text{inf}} + V_{\text{inf}}. \end{aligned}$$

Therefore, letting h be the communication height function defined by (12) with $\theta = q$, we have $\tilde{h} - V_{\text{inf}} = h - U_{\text{inf}}$, and thus $\tilde{H} = H$, $\tilde{H}_c = H_c$ and $\tilde{D} = D$, where H, H_c and D are, respectively, the height function, the critical depth and the difficulty of (Ω, U, q) . It

follows from Theorem 1 that for any family $\{(\beta_n^{\sigma,K})_{n \in \llbracket 1, \sigma K \rrbracket}\}_{\sigma, K \in \mathbb{N}^*}$ of cooling sequences of the form of (40) with $A > H_c$, $\xi > H_c/D$, and $\eta < \xi$, the family $\{(X_n^{\sigma,K})_{n \in \mathbb{N}^*}\}_{\sigma, K \in \mathbb{N}^*}$ of finite-time algorithms $(X_n^{\sigma,K})_n = \text{SC}(\Omega, (U_\beta), (q_\beta), (\beta_n^{\sigma,K}))$ satisfies

$$\forall \varepsilon > 0, \quad \forall \sigma \geq \frac{\ln(\xi/\eta)}{\ln(1 + \varepsilon)},$$

$$\liminf_{K \rightarrow +\infty} - \frac{\ln \sup_{x \in \Omega} \mathbf{P}(U(X_{\sigma K}^{\sigma,K}) \geq U_{\text{inf}} + \eta \mid X_0^{\sigma,K} = x)}{\ln(\sigma K)} \geq \frac{1}{(1 + \varepsilon)D}. \quad (42)$$

In particular, taking

$$\eta < \inf_{x \in \Omega \setminus \Omega_0(U)} U(x) - U_{\text{inf}}$$

gives an asymptotic bound on the probability of failure $\mathbf{P}(X_{\sigma K}^{\sigma,K} \notin \Omega_0(U))$. This result is stated in Theorem 2. Putting it simply, increasing the length K of the constant-temperature stages makes it possible for the probability of failure to decrease as a power of the inverse horizon $(\sigma K)^{-1}$ arbitrarily close to $1/D$, which is the optimal convergence speed exponent of standard SA (see (14)). We recall that D_M is the ‘‘Metropolis difficulty’’ of (Ω, U, q) defined in (20), and we define the convergence measure $\mathcal{M}(\sigma, K)$ of $(X_n^{\sigma,K})_n$ by

$$\mathcal{M}(\sigma, K) = \sup_{x \in \Omega} \mathbf{P}(X_{\sigma K}^{\sigma,K} \notin \Omega_0(U) \mid X_0^{\sigma,K} = x). \quad (43)$$

Theorem 2 *Let $(\Omega, (U_\beta), (q_\beta))$ be an SC process with target energy landscape (Ω, U, q) and satisfying assumptions (A1)–(A3). For any $\varepsilon > 0$ and for any $\sigma \in \mathbb{N}^*$ such that*

$$\sigma > \frac{\ln(D_M/D)}{\ln(1 + \varepsilon)}, \quad (44)$$

there is a family of finite-time cooling sequences $\{(\beta_n^{\sigma,K})_{n \in \llbracket 1, \sigma K \rrbracket}\}_{K \in \mathbb{N}^}$ such that the family of SC algorithms*

$$\{(X_n^{\sigma,K})_{n \in \llbracket 0, \sigma K \rrbracket} = \text{SC}(\Omega, (U_\beta), (q_\beta), (\beta_n^{\sigma,K})) ; K \in \mathbb{N}^*\}$$

satisfies

$$\lim_{K \rightarrow +\infty} - \frac{\ln \mathcal{M}(\sigma, K)}{\ln(\sigma K)} \geq \frac{1}{(1 + \varepsilon)D}. \quad (45)$$

These cooling sequences are piecewise-constant exponential sequences of the form

$$\beta_n^{\sigma,K} = \frac{\ln K}{A} \exp\left(\frac{B}{\sigma} \left\lfloor \frac{n-1}{K} \right\rfloor\right) \quad (46)$$

$$\text{with } \begin{cases} A > H_c \\ \ln(D_M/D) < B < \sigma \ln(1 + \varepsilon). \end{cases} \quad (47)$$

Interestingly, the assumptions of Theorem 2 do not involve the continuation scheme $(U_\beta)_\beta$ (except for pointwise convergence to the target energy). Moreover, the conditions on the communication scheme $(q_\beta)_\beta$ and its limit q are weak. Assumptions (A1) and (A3) are standard in SA theory: the irreducibility of q and the symmetry of its support ensure that the target energy landscape can be fully explored and that any path in this landscape can be traveled in the opposite direction. Assumption (A2) holds if the following two conditions are satisfied:

- (i) $\forall x \in \Omega, q(x, x) = 0 \implies \exists \alpha > 0, q_\beta(x, x) \geq 1/\beta^\alpha;$
- (ii) $\Delta(q_\beta) = \Delta(q)$ for β large enough.

Therefore, since (i) trivially holds if $q(x, x) > 0$ for all x , it suffices to allow the limit communication mechanism to rest anywhere and to “freeze” the set of possible moves at low temperatures.

Remark 3 If, in addition to the assumptions of Theorem 2, (27) holds and the maps $\beta \mapsto q_\beta(x, y)$ and $\beta \mapsto U_\beta(x)$ are continuous for all (x, y) , then SC is “log-optimal” in the sense that

$$\lim_{K \rightarrow +\infty} -\frac{\ln \mathcal{M}(\sigma, K)}{\ln(\sigma K)} = \frac{1}{D} \tag{48}$$

(the proof is similar to that of Theorem 1 in [48]). Hence it may seem strange that near log-optimality can be obtained without condition on the speed at which the continuation scheme converges to the target energy. The truth is that the slower $(U_\beta)_\beta$ converges to U , the larger the length K of the temperature stages to achieve log-optimal convergence within a given tolerance. More precisely, for any $\varepsilon > 0$, any $\sigma \in \mathbb{N}^*$ satisfying (44), and any tolerance $\varepsilon' > \varepsilon$, the lower the speed of convergence of $(U_\beta)_\beta$, the larger the infimum

$$K(\varepsilon') = \inf \left\{ K \in \mathbb{N}^* \mid -\frac{\ln \mathcal{M}(\sigma, K)}{\ln(\sigma K)} \geq \frac{1}{(1 + \varepsilon')D} \right\}. \tag{49}$$

We can link $K(\varepsilon')$ to the speed of convergence of $(U_\beta)_\beta$ by agreeing that a condition for practical convergence to a global minimum of U is that the final inverse temperature $\beta_{\text{sup}} := \beta_{\sigma K}^{\sigma, K}$ be such that the global minima of U_β are global minima of U for any $\beta \geq \beta_{\text{sup}}$. Let us consider a cooling sequence in the class defined by (46) and (47) and measure the speed of convergence of $(U_\beta)_\beta$ by the minimum inverse temperature β_0 from which the global minima of U_β stay in $\Omega_0(U)$, that is,

$$\beta_0 = \inf \{ \beta^* \geq 0 \mid \forall \beta \geq \beta^*, \Omega_0(U_\beta) \subseteq \Omega_0(U) \} \tag{50}$$

(the finite state-space assumption and the pointwise convergence of $(U_\beta)_\beta$ to U guarantee the existence of this infimum). Then, we should have $\beta_{\text{sup}} \geq \beta_0$, which implies that

$$K \geq \exp\left(\frac{H_c \beta_0}{(1 + \varepsilon)^{\sigma-1}}\right), \tag{51}$$

and thus $K(\varepsilon') \geq C \exp(H_c \beta_0)$, where the constant C is a function of ε and σ . In other words, the minimum length of the temperature stages for near log-optimal convergence increases exponentially with the critical depth of the energy landscape and with β_0 .

3.3 Practical considerations

The generation of a realization of a continuation chain $\text{SC}(\Omega, (U_\beta), (q_\beta), (\beta_n))$ is the same as that of an annealing chain $\text{SA}(\Omega, U, \theta, (\beta_n))$, but with U and θ respectively replaced by U_{β_n} and q_{β_n} . For a piecewise-constant cooling sequence $(\beta_n^{\sigma, K})_{n \in \llbracket 1, \sigma K \rrbracket}$ with σ stages of length K , the construction is the following:

- pick an initial state $x_0 \in \Omega;$
- for** $i = 1$ **to** σ **do**

```

set  $\beta \leftarrow \beta_{(i-1)K+1}^{\sigma,K}$ ;
for  $j = 1$  to  $K$  do
  set  $n \leftarrow (i - 1)K + j$ ;
  draw a state  $y$  from the probability distribution  $q_\beta(x_{n-1}, \cdot)$  on  $\Omega$ ;
  set  $x_n \leftarrow x_{n-1}$ ;
  set  $\delta \leftarrow U_\beta(y) - U_\beta(x_{n-1})$ ;
  if  $\delta \leq 0$  then set  $x_n \leftarrow y$ ;
  else set  $x_n \leftarrow y$  with probability  $\exp(-\beta\delta)$ ;
end(if)
end(for)
end(for)
    
```

The time-complexity of SC is governed by the evaluation of the energy difference $U_{\beta_n}(y) - U_{\beta_n}(x_{n-1})$ that takes place at each iteration. Yet, regarding performance, everything is about smart design of the continuation and communication schemes, which is application-dependent. Although not necessary, the choice of $(U_\beta)_\beta$ and $(q_\beta)_\beta$ can be guided by the objective of keeping the average time-complexity of computing $U_\beta(y) - U_\beta(x)$ for $(x, y) \in \Delta(q_\beta)$ of the same order as that of computing $U(y) - U(x)$ for $(x, y) \in \Delta(q)$. In this case, putting aside possible updating operations at the beginning of each constant-temperature stage, SC with piecewise-constant cooling has the same time-complexity as SA.

The performance of an SC algorithm is also strongly influenced by the tuning of its cooling schedule. In most practical situations, it is not possible to obtain good estimates of the bounds on the constants A and B in the cooling sequence (46) suggested by theory—at least not within a reasonable amount of computation time. Generally, the number σ of constant-temperature stages is fixed in advance and the length K of each stage is fixed by the available computing resources. We can then rewrite (46) as

$$\beta_n^{\sigma,K} = \beta_{\text{inf}} \left(\frac{\beta_{\text{sup}}}{\beta_{\text{inf}}} \right)^{\frac{1}{\sigma-1} \lfloor \frac{n-1}{K} \rfloor}, \tag{52}$$

which leaves us with the problem of finding appropriate values for the initial and final inverse temperatures β_{inf} and β_{sup} . To do so, we can consider the following heuristics borrowed from the practice of SA.

1. Most transitions should be accepted at the beginning of the optimization process; that is, letting $(X_n)_n$ be the homogeneous Markov chain with transition matrix $Q_{\beta_{\text{inf}}}$, the acceptance rate

$$\sum_{x \in \Omega} v_{\beta_{\text{inf}}}(x) \sum_{y \in \Omega \setminus \{x\}} Q_{\beta_{\text{inf}}}(x, y) = \text{plim}_{M \rightarrow +\infty} \frac{1}{M} \sum_{n=1}^M \mathbb{1}_{\{X_n \neq X_{n-1}\}}$$

should be close to one. (The probability limit exists if $Q_{\beta_{\text{inf}}}$ is irreducible.)

2. If a global minimum x^\dagger of the target energy U is reached by the end of the optimization process, then the probability to leave x^\dagger by moving uphill should be negligible; that is,

$$\sum_{z \in \Omega : U(z) > U(x^\dagger)} Q_{\beta_{\text{sup}}}(x^\dagger, z)$$

should be close to zero. Of course, in practice, we don't know any ground state of U , and x^\dagger is replaced with a local minimum computed deterministically.

Accurate methods to estimate β_{inf} and β_{sup} according to the above criteria can be found in [51], but they are time-consuming. Besides, high accuracy is usually unnecessary, as exponential cooling is not greatly affected by excessively high initial temperatures or by excessively low final temperatures. The truth is that, as long as the horizon σK is large enough, correct orders of magnitude are satisfactory and hence fast approximate estimation methods are sufficient. In this spirit, we propose to select β_{inf} and β_{sup} so that the uphill acceptance rates (that is, the ratios of the number of accepted uphill moves to the number of proposed ones) at the beginning and at the end of the optimization process are close to some given values $\chi_{\beta_{\text{inf}}}$ and $\chi_{\beta_{\text{sup}}}$. For this purpose, the initial energy landscape $(\Omega, U_{\beta_{\text{inf}}}, q_{\beta_{\text{inf}}})$ is approximated by the infinite-temperature energy landscape (Ω, U_0, q_0) , and the final energy landscape $(\Omega, U_{\beta_{\text{sup}}}, q_{\beta_{\text{sup}}})$ is approximated by the target energy landscape (Ω, U, q) . The procedures are the following; they are used in our experiments in Sect. 4.

1. The estimation of β_{inf} uses the Markov chain $(X_n)_n$ defined by the communication matrix q_0 : given $M \in \mathbb{N}^*$, generate a finite-time realization $(x_n)_n$ of $(X_n)_n$ with exactly M uphill moves with respect to U_0 (that is, M pairs (x_{n_k}, x_{n_k+1}) of successive states such that $U_0(x_{n_k}) < U_0(x_{n_k+1})$, $k \in \llbracket 1, M \rrbracket$), and set β_{inf} to be the solution of

$$\sum_{k=1}^M \exp(-\beta(U_0(x_{n_k+1}) - U_0(x_{n_k}))) = M\chi_{\beta_{\text{inf}}}, \tag{53}$$

which can be determined by any standard root-finding method.

2. Similarly, β_{sup} is estimated from a realization $(y_n)_n$ of the Markov chain with transition matrix q by considering the M first uphill moves (y_{n_k}, y_{n_k+1}) with respect to the target energy U ; that is, β_{sup} is set to be the solution of

$$\sum_{k=1}^M \exp(-\beta(U(y_{n_k+1}) - U(y_{n_k}))) = M\chi_{\beta_{\text{sup}}}. \tag{54}$$

Note that looking only for estimates with correct order of magnitude gives some latitude in choosing $\chi_{\beta_{\text{inf}}}$ and $\chi_{\beta_{\text{sup}}}$. From our own experience, taking $\chi_{\beta_{\text{inf}}} \in [0.7, 0.9]$ and $\chi_{\beta_{\text{sup}}} \in [10^{-4}, 10^{-3}]$ gives exponential cooling schedules with similar performance independently of the application (provided the horizon σK is sufficiently large). The number M of considered uphill moves can be set in accordance with the size of the optimization problem; for instance, choosing M of the order of $100d$ is suitable for the case where Ω is a d -dimensional product space included in \mathbb{R}^d .

4 Application to graph drawing

This section is devoted to the application of SC to the optimization issue associated with the graph-embedding approach of Kamada and Kawai [34]. The problem is formulated in Sect. 4.1 and the proposed SC algorithm is described in Sect. 4.2. Experimental results are given in Sect. 4.3.

4.1 Formal statement of the problem

Let $G = (\mathcal{V}, \mathcal{E})$ be an undirected, connected graph with vertex set \mathcal{V} and edge set $\mathcal{E} \subseteq C_2^{\mathcal{V}}$, where $C_2^{\mathcal{V}}$ denotes the set of 2-subsets of \mathcal{V} . Let $\mathcal{L} = \llbracket 1, L \rrbracket^2$ be a 2-D square lattice of size $L \gg |\mathcal{V}|$, and let $\Omega = \mathcal{L}^{\mathcal{V}}$ be the set of maps from \mathcal{V} to \mathcal{L} , that is, the set of straight-line

drawings of G on \mathcal{L} . The objective is to find a global minimum of the energy $U : \Omega \rightarrow \mathbb{R}$ defined by

$$U(\varphi) = \sum_{\{u,v\} \in \mathcal{C}_2^{\mathcal{V}}} \left(\frac{\|\varphi(u) - \varphi(v)\|}{d_G(u, v)} - \lambda \right)^2, \tag{55}$$

where $\|\cdot\|$ is the standard Euclidean norm, $d_G(u, v)$ is the graph-theoretic distance between vertices u and v (that is, the length of the shortest paths in G between u and v), and the constant λ is the ideal length of an edge in the drawing. A typical choice for λ is

$$\lambda = \frac{L + 1}{\text{diam}(G)}, \tag{56}$$

where $\text{diam}(G)$ stands for the diameter of G .

The above energy U is a particular case of the raw stress criterion associated with least-squares multidimensional scaling (see, for instance, [5]). Indeed, given an arbitrary permutation $(v_1, \dots, v_{|\mathcal{V}|})$ of \mathcal{V} , we have

$$U(\varphi) = \sum_{1 \leq i < j \leq |\mathcal{V}|} w_{ij} (\lambda_{ij} - \delta_{ij}(\varphi))^2, \tag{57}$$

where $w_{ij} = 1/d_G^2(v_i, v_j)$, $\lambda_{ij} = \lambda d_G(v_i, v_j)$, and $\delta_{ij}(\varphi) = \|\varphi(v_i) - \varphi(v_j)\|$. From [26, 55], we know that the extension of (57) to $(\mathbb{R}^p)^{\mathcal{V}}$ generally has non-global minima and that the associated local minima issue is severe when the dimensionality p is small, as it is in our case.

4.2 Specification of the algorithms

Here, we describe our choices for the annealing algorithm $SA(\Omega, U, \theta, (\beta_n))$ and for the continuation algorithm $SC(\Omega, (U_\beta), (q_\beta), (\beta_n))$ used in our experiments. We start with the communication matrix θ of the annealing algorithm and we continue with the description of the continuation scheme $(U_\beta)_\beta$. Then, we discuss the design of the communication scheme $(q_\beta)_\beta$, whose definition involves θ and whose choice is implicitly guided by that of $(U_\beta)_\beta$. We end this section with the specification of the functions controlling the continuation and communication schemes.

4.2.1 The annealing communication matrix

The design of a communication mechanism for SA must balance two conflicting objectives. On the one hand, the annealing chain should mix rapidly at high temperatures to ensure efficient exploration of the state space, which calls for large communication neighborhoods. On the other hand, it is desirable that the acceptance rate remains significant in the low temperature regime, which requires small communication neighborhoods. Another way of seeing things is that the larger the set of possible moves, the less numerous and the better the local minima of the energy landscape, but the less mobile the annealing chain at mid and low temperatures.

We propose to consider a single-vertex updating dynamics in which the amplitude of the vertex moves obeys a Rayleigh distribution. More specifically, the neighbors of a given layout φ are the layouts that differ from φ by the location of a single vertex, and the difference vector between the candidate and the current positions of a vertex is of the form

$$\delta(r, \omega) = ([r \cos \omega], [r \sin \omega]), \tag{58}$$

where $[\cdot]$ denotes the nearest integer function, the angle ω is a sample from the uniform density on $[0, 2\pi]$, and the radius r is a sample from the density

$$f_R : r \in \mathbb{R}_+ \mapsto \frac{\pi r}{2R^2} \exp\left(\frac{-\pi r^2}{4R^2}\right), \tag{59}$$

where the mean value R is fixed. The balance between rapid mixing at high temperatures and reasonable mobility at low temperatures is achieved by adjusting R . However, the optimal value R^\dagger of R cannot be predicted in advance because of its intricate dependance on the graph G to be drawn (R^\dagger also depends on the lattice size L and on the ideal edge-length λ). In fact, the only way to estimate R^\dagger is by trial and error, which is prohibitive in terms of computation time.

In practice, the generation of a new candidate solution ψ from φ is performed as follows:

```

set  $\psi \leftarrow \varphi$ ;
select a vertex  $v \in \mathcal{V}$  uniformly at random;
draw  $r$  from  $f_R$ ;
draw  $\omega$  from the uniform density on  $[0, 2\pi]$ ;
if  $\varphi(v) + \delta(r, \omega) \in \mathcal{L}$  then
    set  $\psi(v) \leftarrow \varphi(v) + \delta(r, \omega)$ ;
end(if)
    
```

The corresponding communication matrix is clearly symmetric and irreducible; we denote it by θ_R . Formally, θ_R is of the form of (3) with $c = |\mathcal{V}|$ and with neighborhood system and weighting function respectively given by

$$\mathcal{N}(\varphi) = \{\psi \in \Omega \mid \exists! v \in \mathcal{V}, \psi(v) \neq \varphi(v)\} \tag{60}$$

$$\text{and } \Gamma(\{\varphi, \psi\}) = \int_{S((\psi-\varphi)(v))} \frac{f_R(\|s\|)}{\|s\|} ds, \tag{61}$$

where $S(\mathbf{a})$ denotes the unit square in \mathbb{R}^2 centered at point \mathbf{a} , and v is the (unique) vertex at which φ and ψ differ.

4.2.2 The continuation scheme

We define the continuation scheme $(U_\beta : \Omega \rightarrow \mathbb{R})_\beta$ from the target energy U (55) by simply making the ideal edge-length vary with β , that is,

$$U_\beta(\varphi) = \sum_{\{u,v\} \in \mathcal{C}_2^\mathcal{V}} \left(\frac{\|\varphi(u) - \varphi(v)\|}{d_G(u, v)} - \Lambda(\beta) \right)^2, \tag{62}$$

where the control function $\Lambda : \mathbb{R}_+ \rightarrow (0, \lambda]$ increases monotonically to λ . The main feature of this scheme is the implicit control of the amplitude of candidate vertex moves. Indeed,

$$U_\beta(\varphi) \propto U^* \left(\frac{\lambda}{\Lambda(\beta)} \varphi \right), \tag{63}$$

where U^* is the extension of U to $(\mathbb{R}^2)^\mathcal{V}$. Therefore, moving a vertex location by a distance r when considering U_β is equivalent to a move of length $\lambda r / \Lambda(\beta)$ from the standpoint of the target energy. In particular, using (62) together with the communication mechanism θ_R

defined in Sect. 4.2.1 amounts to considering the target energy and replacing the mean R of the proposed displacement amplitudes by the “effective mean”

$$R_\beta = \frac{\lambda R}{\Lambda(\beta)}. \tag{64}$$

4.2.3 The communication scheme

Intuitively, the communication scheme $(q_\beta : \Omega^2 \rightarrow [0, 1])_\beta$ should allow balanced exploration of the state space at the beginning of the SC process, and it should favor moves towards nearby minima by the end of the SC process. A simple and efficient way to get this behavior is to design two communication matrices \bar{q} and \underline{q} that are respectively adapted to the high- and low-temperature regimes, and to control the probability of choosing one over the other as a function of β ; that is,

$$q_\beta = (1 - \Xi(\beta))\bar{q} + \Xi(\beta)\underline{q}, \tag{65}$$

where $\Xi(\beta)$ is the probability of choosing \underline{q} rather than \bar{q} to generate a new candidate solution. (An alternative is to design different communication matrices and to select the current communication matrix adaptively [2], but this approach does not fit into the SC framework and does not come with a convergence guarantee.) We impose that the control function $\Xi : \mathbb{R}_+ \rightarrow [0, 1]$ is monotonically increasing and that $\Xi_{\text{sup}} := \lim_{\beta \rightarrow +\infty} \Xi(\beta) < 1$. It is then easily seen that assumptions (A1)–(A3) in Propositions 2 and 3 hold—and hence that Theorem 2 applies—if the following conditions are satisfied: (i) \bar{q} is irreducible, (ii) $\bar{q}(x, x) > 0$ for all x , (iii) $\Delta(\bar{q})$ is symmetric, and (iv) $\Delta(\underline{q}) \subseteq \Delta(\bar{q})$.

A natural choice for \bar{q} is the site-by-site communication mechanism θ_R defined in Sect. 4.2.1 with R large enough to ensure rapid mixing at high temperatures. Note that care must be taken when using the continuation scheme (62), since in this case we have to think in terms of the effective mean R_β (64); in our experiments, R is chosen so that the proposed vertex move at the beginning of the SC process have a mean amplitude of $L/10$ (as discussed in the next section).

Concerning the choice of \underline{q} , the only constraint is $\Delta(\underline{q}) \subseteq \Delta(\theta_R)$, and hence any site-by-site communication strategy is appropriate. We propose to use a mechanism similar to θ_R , except that the angle ω is computed deterministically so as to move towards a minimum of the 2-D function obtained by viewing the target energy U as a function of $\varphi(v)$ only. More precisely, given the current solution φ together with the vertex v subject to potential move, we consider the function

$$W_{\varphi,v} : \mathbf{a} \in \mathbb{R}^2 \mapsto U^*(\varphi_{v,\mathbf{a}}), \tag{66}$$

where $\varphi_{v,\mathbf{a}} : \mathcal{V} \rightarrow \mathbb{R}^2$ is given by

$$\varphi_{v,\mathbf{a}}(u) = \begin{cases} \mathbf{a} & \text{if } u = v \\ \varphi(u) & \text{if } u \neq v. \end{cases} \tag{67}$$

Drawing inspiration from the vertex repositioning method of Kamada and Kawai [34], we approach a minimum of $W_{\varphi,v}$ by following the Newton search direction $\delta_N(\varphi, v)$ defined by

$$\nabla^2 W_{\varphi,v}(\varphi(v))\delta_N(\varphi, v) = -\nabla W_{\varphi,v}(\varphi(v)), \tag{68}$$

where $\nabla^2 W_{\varphi,v}$ and $\nabla W_{\varphi,v}$ respectively stand for the Hessian matrix and the column gradient-vector of $W_{\varphi,v}$.

To sum up, then, a candidate layout ψ is drawn from the probability distribution $q_\beta(\varphi, \cdot)$ on Ω as follows:

```

set  $\psi \leftarrow \varphi$ ;
select a vertex  $v \in \mathcal{V}$  uniformly at random;
draw  $r$  from  $f_R$ ;
draw  $p$  from the uniform density on  $[0, 1]$ ;
if  $p \geq \Xi(\beta)$  then
    draw  $\omega$  from the uniform density on  $[0, 2\pi]$ ;
    if  $\varphi(v) + \delta(r, \omega) \in \mathcal{L}$  then
        set  $\psi(v) \leftarrow \varphi(v) + \delta(r, \omega)$ ;
    end(if)
elseif  $\det \nabla^2 W_{\varphi, v}(\varphi(v)) \neq 0$  and  $\nabla W_{\varphi, v}(\varphi(v)) \neq 0$  then
    set  $(a_1, a_2) \leftarrow r \delta_N(\varphi, v) / \|\delta_N(\varphi, v)\|$ ;
    set  $\mathbf{a} \leftarrow ([a_1], [a_2])$ ;
    if  $\varphi(v) + \mathbf{a} \in \mathcal{L}$  then
        set  $\psi(v) \leftarrow \varphi(v) + \mathbf{a}$ ;
    end(if)
end(if)
    
```

4.2.4 The control functions

We simply take the control function Λ of the continuation scheme to increase linearly with the inverse temperature: given $\lambda_{\text{inf}} \in (0, \lambda)$, we set

$$\Lambda(\beta) = \lambda_{\text{inf}} + (\lambda - \lambda_{\text{inf}}) \frac{\beta - \beta_{\text{inf}}}{\beta_{\text{sup}} - \beta_{\text{inf}}}, \tag{69}$$

where β_{inf} and β_{sup} are the initial and final inverse temperatures of the cooling sequence (52). Therefore, according to (64), the mean amplitude R_β of the proposed vertex moves decreases inversely linearly with β . We choose the parameter R of the high-temperature communication matrix ($\bar{q} = \theta_R$ in (65)) so that $R_{\beta_{\text{inf}}}$ is equal to $L/10$, that is,

$$R = \frac{L}{10} \frac{\lambda_{\text{inf}}}{\lambda}. \tag{70}$$

Besides, since $R_{\beta_{\text{sup}}} = R$, the mean amplitude of the proposed vertex moves at the end of the SC process can be freely chosen by adjusting λ_{inf} ; we set $\lambda_{\text{inf}} = \lambda/10$ so that $R_{\beta_{\text{sup}}} = L/100$. For a cooling sequence of the form of (52), the resulting expression of R_β as a function of the iteration index n depends on the initial to final temperature ratio $\beta_{\text{sup}}/\beta_{\text{inf}}$, which in turn depends on the graph to be drawn. For instance, for the test graphs used in our experiments, $\beta_{\text{sup}}/\beta_{\text{inf}}$ lies between 10^3 and 10^4 and the mappings $n \mapsto R_{\beta_n}$ associated with these two values are shown in Fig. 2a (we took $(\sigma, K) = (N, 1)$ to simplify the representations). It should be stressed that our SC algorithm is not sensitive to the choice of Λ as long as R_{β_n} decreases smoothly with n . For example, the performance is the same when R_{β_n} decreases linearly with n or when R_β decreases linearly with increasing temperature. We will not discuss this aspect further.

The choice of the control function Ξ of the communication scheme is slightly more delicate, since encouraging deterministic moves towards local minima too early in the optimization process is a waste of computation time. In order to delay the impact of the semi-deterministic communication mechanism q , we propose to take

Fig. 2 **a** Effective mean of the length of the proposed vertex moves and **b** control function of the communication scheme as functions of the iteration index for $\beta_{\text{sup}}/\beta_{\text{inf}} = 10^3$ (dashed lines) and $\beta_{\text{sup}}/\beta_{\text{inf}} = 10^4$ (solid lines)

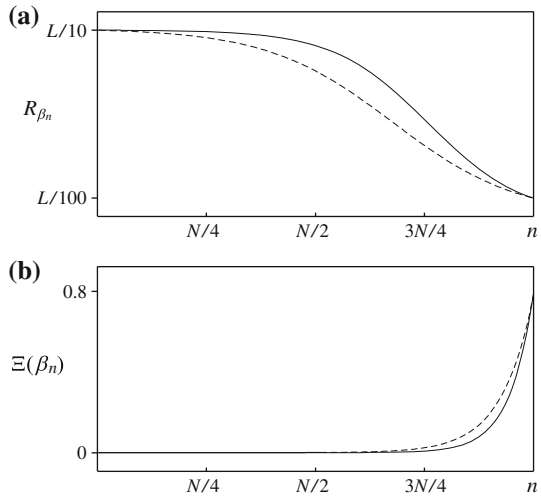
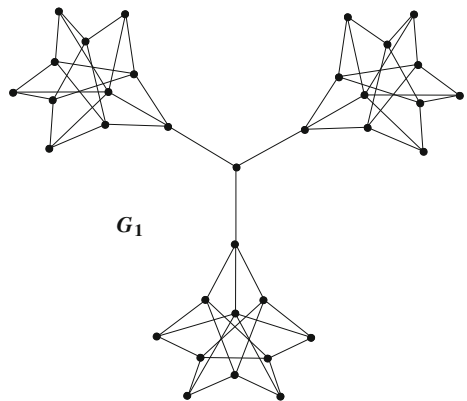


Fig. 3 Graph G_1 : replicas of the Grötzsch graph connected via the leaves of the 3-star graph



$$\Xi(\beta) = \Xi_{\text{sup}} \frac{\beta^2 - (\beta_{\text{inf}})^2}{(\beta_{\text{sup}})^2 - (\beta_{\text{inf}})^2} \tag{71}$$

with $\Xi_{\text{sup}} \in (0, 1)$. The parameter Ξ_{sup} is set to 0.8 in our experiments, but any value greater than $\frac{1}{2}$ is good. The mappings $n \mapsto \Xi(\beta_n)$ corresponding to $\beta_{\text{sup}}/\beta_{\text{inf}} = 10^3$ and 10^4 are shown in Fig. 2b

4.3 Experimental results

To investigate the behavior of the SA and SC algorithms described in Sect. 4.2, we purposely selected five test graphs that challenge both the Kamada–Kawai algorithm and SA. These connected, undirected graphs are shown in Figs. 3, 4 and 5 (the layouts were produced by our SC algorithm). Graph G_1 is obtained by duplicating the Grötzsch graph two times and by merging a degree-3 vertex in each of the three resulting identical graphs with a leave of the claw graph. Graphs G_2 – G_5 are from the 12th and 13th graph drawing contests held in conjunction with the 2005 and 2006 International Symposiums on Graph Drawing. These test graphs have varying number of vertices, varying average vertex-degree and varying diameter; their basic characteristics are summarized in Table 1.

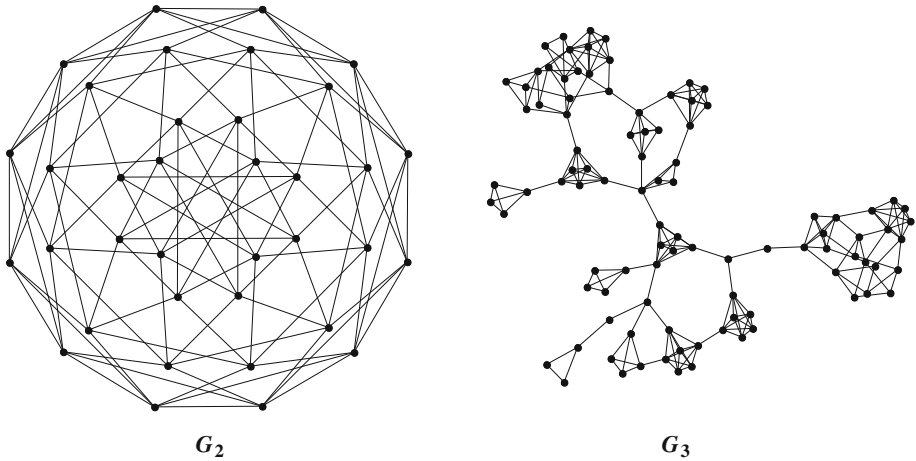


Fig. 4 Graphs G_2 and G_3 : graphs 4 and 1 of the 12th graph drawing contest held in conjunction with the 2005 graph drawing symposium in Limerick, Ireland

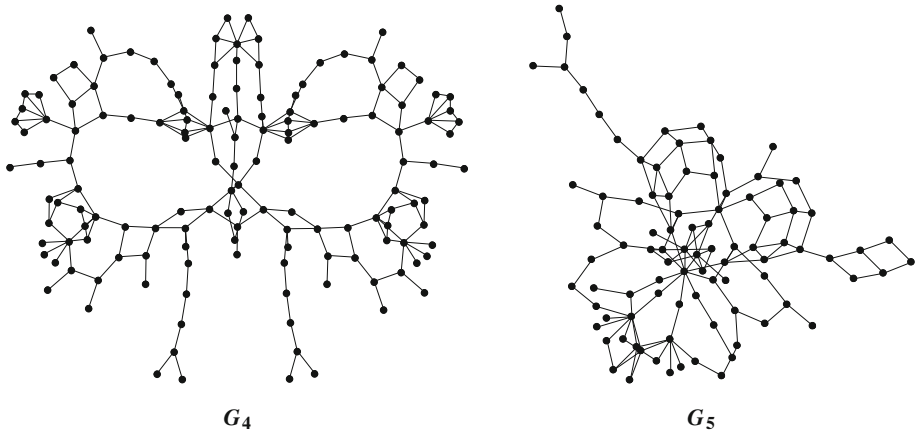


Fig. 5 Graphs G_4 and G_5 : graphs 4 and 2 of the 13th graph drawing contest held in conjunction with the 2006 graph drawing symposium in Karlsruhe, Germany

Table 1 Number of vertices, number of edges, average vertex-degree and diameter of the test graphs displayed in Figs. 3, 4 and 5

$G_i = (\mathcal{V}_i, \mathcal{E}_i)$	G_1	G_2	G_3	G_4	G_5
$ \mathcal{V}_i $	34	36	97	143	92
$ \mathcal{E}_i $	63	108	223	194	133
$2 \mathcal{E}_i / \mathcal{V}_i $	3.71	6	4.60	2.71	2.89
$\text{diam}(G_i)$	6	4	13	18	16

We consider the following algorithms:

1. the enhanced version of the Kamada–Kawai algorithm—denoted by KK—mentioned in [6] and described in [29, §4.2];

2. the annealing algorithm $SA_R := SA(\Omega, U, \theta_R, (\beta_n))$ with $R = L/10$ and with $R = L/100$;
3. the SC algorithm with temperature-dependent energy and fixed communication $SC_1 := SC(\Omega, (U_\beta), \theta_{L/100}, (\beta_n))$;
4. the SC algorithm with temperature-dependent energy and temperature-dependent communication $SC_2 := SC(\Omega, (U_\beta), (q_\beta), (\beta_n))$.

The values $L/10$ and $L/100$ of the parameter R of the SA algorithm coincide with those of the mean amplitude of the proposed vertex moves at the beginning and at the end of the SC algorithms—from our experience, the performance of SA decreases when R lies outside of the range $[L/100, L/10]$. In all cases, the lattice size L is set to 10^3 and the ideal edge-length λ is defined as in (56). The cooling sequence of the SA and SC algorithms is of the form of (52) with horizon $\sigma K = 5 \cdot 10^3 |\mathcal{V}|$ divided into $\sigma = 250$ constant-temperature stages of length $K = 20|\mathcal{V}|$. The initial and final inverse temperatures β_{inf} and β_{sup} are estimated by means of the fast approximate methods proposed in Sect. 3.3 with $\chi_{\beta_{\text{inf}}} = 0.8$, $\chi_{\beta_{\text{sup}}} = 5 \cdot 10^{-4}$, and $M = 100|\mathcal{V}|$.

For each graph $G_i, i \in \llbracket 1, 5 \rrbracket$, each algorithm is run 100 times starting from the elements of a same set of randomly generated layouts. The performance is measured by the averages of the Kamada–Kawai energy (55) and of the maximum-to-minimum edge-length ratio over the drawings $\varphi_1, \dots, \varphi_{100}$ produced by the different runs, that is, by

$$\bar{U} = 10^{-2} \sum_{i=1}^{100} U(\varphi_i) \tag{72}$$

$$\text{and } \bar{q} = 10^{-2} \sum_{i=1}^{100} \frac{\max_{\{u,v\} \in \mathcal{E}} \|\varphi_i(u) - \varphi_i(v)\|}{\min_{\{u,v\} \in \mathcal{E}} \|\varphi_i(u) - \varphi_i(v)\|}. \tag{73}$$

We also consider the number of edge-crossings, which is an important aesthetic criterion for graph layout, although the Kamada–Kawai energy sacrifices it for symmetry. Edge-length dispersion, on the other hand, is left aside because it was not found to be discriminating. Obviously, the smaller \bar{U}, \bar{q} and the average number of edge-crossings, the better.

Before proceeding with the discussion of our results, let us remark that the quality of the layouts produced by the annealing algorithm SA_R is sensitive to the choice of the parameter R . This is exemplified in Fig. 6, which displays the average final energy level \bar{U} obtained with SA_R as a function of R for the graphs G_1 and G_2 . Consequently, for proper comparison, we also include the best results obtainable by SA, that is, those produced by the SA algorithm with the optimal value R^\dagger of R ($R^\dagger = 26, 39, 71, 84$ and 69 for G_1, G_2, G_3, G_4 and G_5 , respectively). However, it is important to be clear that the estimation of R^\dagger is unrealistic in practice. Indeed, the function $R \in (0, L) \mapsto \bar{U}(R)$ associated with SA_R has local valleys and hence, since the computational cost of a single evaluation of \bar{U} is that of annealing, the task of estimating R^\dagger is prohibitive in terms of computation time.

Our results are summarized in Table 2, which gives the averages of the final energy level, the maximum-to-minimum edge-length ratio and the number of edge-crossings for each algorithm and each test graph. Note that the KK algorithm does not always converges and that the associated average values in Table 2 are computed from the outputs of the convergent runs only (KK oscillates in 2% of the cases for G_1 and G_3 , and in 4, 6 and 5% of the cases for G_2, G_4 and G_5 , respectively). We make the following observations.

Fig. 6 Average final energy level over 100 runs of SA_R as a function of R for the graphs G_1 and G_2

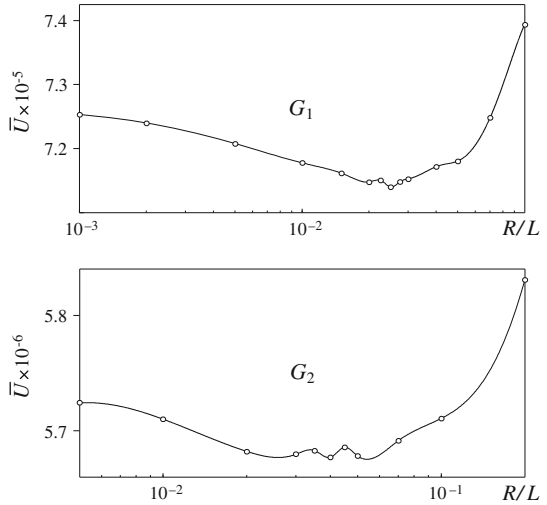


Table 2 Average results over 100 runs of the competing algorithms: Kamada–Kawai (KK), SA with mean vertex-move amplitude equal to $L/10$ and $L/100$ ($SA_{L/10}$ and $SA_{L/100}$), SA with optimal mean vertex-move amplitude (SA_{R^\dagger}), SC with temperature-dependent energy and fixed communication (SC_1), and SC with temperature-dependent energy and temperature-dependent communication (SC_2). Given are the averages of the final energy level, the max-to-min edge-length ratio and the number of edge-crossings for each test graph

Algorithm	G_1	G_2	G_3	G_4	G_5
Average final energy level $\bar{U}(\times 10^{-5})$					
KK	7.311	61.429	9.654	11.388	10.008
$SA_{L/10}$	7.394	57.107	8.942	9.573	9.636
$SA_{L/100}$	7.178	57.100	9.510	11.447	9.771
SA_{R^\dagger}	7.140	56.771	8.736	9.403	9.547
SC_1	7.144	56.785	8.700	9.306	9.520
SC_2	7.130	56.716	8.606	9.086	9.427
Average max-to-min edge-length ratio $\bar{\rho}$					
KK	2.55	4.30	4.30	3.85	2.93
$SA_{L/10}$	3.07	3.18	6.37	4.81	4.10
$SA_{L/100}$	2.43	2.83	3.96	3.51	2.85
SA_{R^\dagger}	2.50	2.66	4.99	4.77	3.41
SC_1	2.51	2.64	3.90	3.83	2.88
SC_2	2.46	2.55	3.67	3.73	2.76
Average number of edge-crossings					
KK	54	252	136	18	46
$SA_{L/10}$	50	209	128	21	40
$SA_{L/100}$	52	237	132	15	41
SA_{R^\dagger}	51	212	127	15	38
SC_1	50	211	128	14	40
SC_2	50	209	127	14	38

Final energy level

Predictably enough, the optimally tuned annealing algorithm SA_{R^\dagger} outperforms KK for all test graphs. However, the performance of SA can vary significantly with the mean vertex-move amplitude R (as is the case for G_3 and G_4) and SA can even perform worse than the KK algorithm for plausible values of R (as happens for G_1 with $R = L/10$ and for G_4 with $R = L/100$). The SC algorithm SC_1 obtained by making the energy in $SA_{L/100}$ temperature-dependent performs similarly to SA_{R^\dagger} ; in other words, the proposed continuation scheme—simple though it is—allows to circumvent the problem of selecting an appropriate value for R . Moreover, the introduction of the temperature-dependent communication mechanism further improves performance: the average energy level achieved by the full SC algorithm SC_2 is lower than those associated with SC_1 and SA_{R^\dagger} in all cases.

Maximum-to-minimum edge-length ratio

The average maximum-to-minimum edge-length ratio of the layouts produced by SA depends strongly on R for all test graphs; it tends to decrease with increasing R because large amplitude moves preclude fine edge-length adjustments. Quite surprisingly, SA_{R^\dagger} does not always perform better than KK according to this aesthetic criterion, as observed for G_3 , G_4 and G_5 . Since SA_{R^\dagger} outperforms KK in terms of minimization of the Kamada–Kawai energy, this shows that there exists low-energy drawings with unbalanced edge-length, and it is tempting to conclude that the Kamada–Kawai energy does not control edge-length dispersion. But the results produced by SC do not support this claim. Indeed, in our experiments, we observed that SC systematically outperforms SA_{R^\dagger} in terms of maximum-to-minimum edge-length ratio, and that the improvement brought by SC tends to increase with the number of vertices. For the examples considered here, the percentage decrease of $\bar{\varrho}$ when using SC_2 over SA_{R^\dagger} ranges from 1.6% for graph G_1 to 26% for graph G_3 .

Number of edge-crossings

Depending on the graph structure, and just as for the two other performance measures, the average number of edge-crossing of the layouts produced by SA can vary significantly with R , as observed for G_2 and G_4 . The drawings produced by SA_{R^\dagger} have a smaller number of edge-crossings than those generated by KK (the corresponding average percentage decrease ranges from 5.5% for graph G_1 to 17% for graph G_5), but SA_{R^\dagger} performs slightly worse than $SA_{L/10}$ in the case of G_1 and G_2 . The full continuation algorithm SC_2 outperforms all three annealing algorithms in any case.

Summary

In summary, SC_2 systematically outperforms the optimally tuned annealing algorithm SA_{R^\dagger} in terms of three standard aesthetic criteria (the Kamada–Kawai energy, the maximum-to-minimum edge-length ratio and the number of edge-crossings), and SC_2 always performs better than $SA_{L/10}$ and than $SA_{L/100}$ in terms of the Kamada–Kawai energy. It may happen that annealing produces better results than SC_2 according to either edge-length dispersion or edge-crossings, but not both (for instance, in the presented experiments, the layouts of G_1 and G_4 produced by $SA_{L/100}$ have smaller edge-length dispersion than those generated by SC_2). However, the performance difference in question is usually not significant and is always balanced by the the two other aesthetic criteria.

5 Conclusion

A natural generalization of SA is to allow the communication mechanism and the energy function to be temperature-dependent in an attempt to speed up the optimization process; we call the class of such algorithms SC. We derived simple sufficient conditions for the global convergence of SC that drastically increase the flexibility in designing annealing-type algorithms: the assumptions on the communication mechanism are weak, and, even more interestingly, there is no condition on the speed of convergence of the energy towards the target objective function. Besides, simple exponential cooling sequences with constant-temperature stages of constant length make it possible for the convergence speed exponent of SC to be arbitrarily close to the best exponent of SA over all possible cooling sequences.

On the practical side, good SC algorithm design means appropriate tuning of the cooling schedule and clever temperature-parameterization of the communication mechanism and of the energy function. We provided efficient generic methods for selecting the initial and final temperatures of the cooling schedule, and we demonstrated the benefits of SC via an original application to the optimization issue associated with the graph drawing approach of Kamada and Kawai [34]. Our experimental results led to the conclusion that well-designed SC algorithms can outperform SA with optimal communication setting without requiring additional computational efforts. More generally, the flexibility and the favorable convergence properties of SC makes it potentially attractive for a wide range of difficult optimization problems.

Acknowledgments This work was partly supported by the French National Research Agency under grant ANR-09-BLAN-0372-01

References

1. Aarts, E.H.L., Korst, J.M., van Laarhoven, P.J.M.: A quantitative analysis of the simulated annealing algorithm: a case study for the traveling salesman problem. *J. Stat. Phys.* **50**(1/2), 187–206 (1988)
2. Alizamir, S., Rebennack, S., Pardalos, P.: Improving the neighborhood selection strategy in simulated annealing using optimal stopping problem. In: Tan, C. (ed.) *Simulated Annealing*, pp. 363–382. I-Tech Education and Publishing, Austria (2008)
3. Bélisle, C.: Convergence theorems for a class of simulated annealing algorithms on \mathbb{R}^d . *J. Appl. Probab.* **29**(4), 885–895 (1992)
4. Blake, A., Zisserman, A.: *Visual Reconstruction*. MIT Press, Cambridge (1987)
5. Borg, I., Groenen, P.: *Modern Multidimensional Scaling*. Springer, Berlin (2009)
6. Brandenburg, F., Himsolt, M., Rohrer, C.: An experimental comparison of force-directed and randomized graph drawing algorithms. In: *Proceedings of the 1995 Symposium on Graph Drawing*, Lecture Notes in Computer Science, vol. 1027, pp. 76–87. Springer, Berlin (1996)
7. Bryan, K., Cunningham, P., Bolshakova, N.: Application of simulated annealing to the biclustering of gene expression data. *IEEE Trans. Inf. Technol. Biomed.* **10**(3), 519–525 (2006)
8. Catoni, O.: Large deviations and cooling schedules for the simulated annealing algorithm (in french). *C. R. Acad. Sci. Paris Sér. I Math.* **307**, 535–538 (1988)
9. Catoni, O.: Rough large deviation estimates for simulated annealing: application to exponential schedules. *Ann. Probab.* **20**(3), 1109–1146 (1992)
10. Catoni, O.: Metropolis, simulated annealing, and iterated energy transformation algorithms: theory and experiments. *J. Complex.* **12**(4), 595–623 (1996)
11. Catoni, O.: Solving scheduling problems by simulated annealing. *SIAM J. Control Optim.* **36**(5), 1539–1575 (1998)
12. Catoni, O.: Simulated annealing algorithms and Markov chains with rare transitions. In: Azema, J., Emery, M., Ledoux, M., Yor, M. (eds) *Séminaire de probabilités XXXIII*, Lecture Notes in Mathematics, vol. 1709, pp. 69–119. Springer, New York (1999)
13. Chiang, T.S., Chow, Y.: On the convergence rate of annealing processes. *SIAM J. Control Optim.* **26**(6), 1455–1470 (1988)

14. Cohn, H., Fielding, M.: Simulated annealing: searching for an optimal temperature schedule. *SIAM J. Optim.* **9**(3), 779–802 (1999)
15. Cot, C., Catoni, O.: Piecewise constant triangular cooling schedules for generalized simulated annealing algorithms. *Ann. Appl. Probab.* **8**(2), 375–396 (1998)
16. Davidson, R., Harel, D.: Drawing graphs nicely using simulated annealing. *ACM Trans. Graph.* **15**(4), 301–331 (1996)
17. de Amorim, S., Barthélemy, J.P., Ribeiro, C.: Clustering and clique partitioning: simulated annealing and tabu search approaches. *J. Classif.* **9**(1), 17–41 (1992)
18. Del Moral, P., Miclo, L.: On the convergence and applications of generalized simulated annealing. *SIAM J. Control Optim.* **37**(4), 1222–1250 (1999)
19. Desai, M.: Some results characterizing the finite time behaviour of the simulated annealing algorithm. *Sādhanā* **24**(4–5), 317–337 (1999)
20. Fielding, M.: Simulated annealing with an optimal fixed temperature. *SIAM J. Optim.* **11**(2), 289–307 (2000)
21. Frick, A., Ludwig, A., Mehldau, H.: A fast adaptive layout algorithm for undirected graphs. In: *Proceedings of the 1994 Symposium on Graph Drawing, Lecture Notes in Computer Science*, vol. 894, pp. 388–403. Springer, Berlin (1995)
22. Frigerio, A., Grillo, G.: Simulated annealing with time-dependent energy function. *Math. Z.* **213**, 97–116 (1993)
23. Fruchterman, T.M.J., Reingold, E.M.: Graph drawing by force-directed placement. *Softw. Pract. Exper.* **21**(11), 1129–1164 (1991)
24. Gelfand, S., Mitter, S.: Metropolis-type annealing algorithms for global optimization in \mathbb{R}^d . *SIAM J. Control Optim.* **31**(1), 111–131 (1993)
25. Gidas, B.: Nonstationary Markov chains and convergence of the annealing algorithm. *J. Stat. Phys.* **39**(1/2), 73–131 (1985)
26. Groenen, P., Heiser, W.: The tunneling method for global optimization in multidimensional scaling. *Psychometrika* **61**(3), 529–550 (1996)
27. Haario, H., Saksman, E.: Simulated annealing process in general state space. *Adv. Appl. Probab.* **23**(4), 866–893 (1991)
28. Hajek, B.: Cooling schedules for optimal annealing. *Math. Oper. Res.* **13**(2), 311–329 (1988)
29. Harel, D., Koren, Y.: A fast multi-scale method for drawing large graphs. *J. Graph Algorithms Appl.* **6**(3), 179–202 (2002)
30. Jacobson, S., Yücesan, E.: Analyzing the performance of generalized hill climbing algorithms. *J. Heuristics* **10**(4), 387–405 (2004)
31. Johnson, A., Jacobson, S.: On the convergence of generalized hill climbing algorithms. *Discrete Appl. Math.* **119**(1–2), 37–57 (2002)
32. Johnson, D., Aragon, C., McGeoch, L., Shevon, C.: Optimization by simulated annealing: an experimental evaluation; part I, graph partitioning. *Oper. Res.* **37**(6), 865–892 (1989)
33. Johnson, V., Wong, W., Hu, X., Chen, C.T.: Image restoration using Gibbs priors: boundary modeling, treatment of blurring, and selection of hyperparameters. *IEEE Trans. Pattern Anal. Mach. Intell.* **13**(5), 413–425 (1991)
34. Kamada, T., Kawai, S.: An algorithm for drawing general undirected graphs. *Inf. Process. Lett.* **31**(1), 7–15 (1989)
35. Kaufmann, M., Wagner, D. (eds.): *Drawing Graphs: Methods and Models, Lecture Notes in Computer Science*, vol. 2025. Springer, Berlin (2001)
36. Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P.: Optimization by simulated annealing. *Science* **220**(4598), 671–680 (1983)
37. Locatelli, M.: Convergence properties of simulated annealing for continuous global optimization. *J. Appl. Probab.* **33**(4), 1127–1140 (1996)
38. Locatelli, M.: Simulated annealing algorithms for continuous global optimization: convergence conditions. *J. Optim. Theory Appl.* **104**(1), 121–133 (2000)
39. Löwe, M.: Simulated annealing with time-dependent energy function via Sobolev inequalities. *Stoch. Process. Appl.* **63**(2), 221–233 (1996)
40. Nielsen, M.: Graduated nonconvexity by functional focusing. *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(5), 521–525 (1997)
41. Nikolaev, A., Jacobson, S.: Simulated annealing. In: Gendreau, M., Potvin, J.Y. (eds.) *Handbook of Meta-heuristics, International Series in Operations Research and Management Science*, vol. 146, pp. 1–40. Springer, Berlin (2010)
42. Nikolova, M.: Markovian reconstruction using a GNC approach. *IEEE Trans. Image Process.* **8**(9), 1204–1220 (1999)

43. Ohlmann, J., Bean, J., Henderson, S.: Convergence in probability of compressed annealing. *Math. Oper. Res.* **29**(4), 837–860 (2004)
44. Orosz, J., Jacobson, S.: Analysis of static simulated annealing algorithms. *J. Optim. Theory Appl.* **115**(1), 165–182 (2002)
45. Orosz, J., Jacobson, S.: Finite-time performance analysis of static simulated annealing algorithms. *Comput. Optim. Appl.* **21**(1), 21–53 (2002)
46. Purchase, H.: Metrics for graph drawing aesthetics. *J. Vis. Lang. Comput.* **13**(5), 501–516 (2002)
47. Robini, M.C., Bresler, Y., Magnin, I.E.: On the convergence of Metropolis-type relaxation and annealing with constraints. *Probab. Eng. Inf. Sci.* **16**(4), 427–452 (2002)
48. Robini, M.C., Lachal, A., Magnin, I.E.: A stochastic continuation approach to piecewise constant reconstruction. *IEEE Trans. Image Process.* **16**(10), 2576–2589 (2007)
49. Robini, M.C., Magnin, I.E.: Stochastic nonlinear image restoration using the wavelet transform. *IEEE Trans. Image Process.* **12**(8), 890–905 (2003)
50. Robini, M.C., Magnin, I.E.: Optimization by stochastic continuation. *SIAM J. Imaging Sci.* **3**(4), 1096–1121 (2010)
51. Robini, M.C., Rastello, T., Magnin, I.E.: Simulated annealing, acceleration techniques and image restoration. *IEEE Trans. Image Process.* **8**(10), 1374–1387 (1999)
52. Robini, M.C., Reissman, P.J.: On simulated annealing with temperature-dependent energy and temperature-dependent communication. *Stat. Probab. Lett.* **81**(8), 915–920 (2011)
53. Schuur, P.: Classification of acceptance criteria for the simulated annealing algorithm. *Math. Oper. Res.* **22**(2), 266–275 (1997)
54. Tollis, I., Battista, G.D., Eades, P., Tamassia, R.: *Graph Drawing: Algorithms for the Visualization of Graphs*. Prentice Hall, Englewood Cliffs, NJ (1999)
55. Trosset, M.: On the existence of nonglobal minimizers of the stress criterion for metric multidimensional scaling. In: *Proceedings of the Statistical Computing Section*, pp. 158–162. ASA (1997)
56. Trounev, A.: Cycle decompositions and simulated annealing. *SIAM J. Control Optim.* **34**(3), 966–986 (1996)
57. Wang, J.P.: Stochastic relaxation on partitions with connected components and its application to image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **20**(6), 619–636 (1998)
58. Ware, C., Purchase, H., Colpoys, L., McGill, M.: Cognitive measurements of graph aesthetics. *Inf. Vis.* **1**(2), 103–110 (2002)
59. Yang, R.: Convergence of the simulated annealing algorithm for continuous global optimization. *J. Optim. Theory Appl.* **104**(3), 691–716 (2000)