Optimization Methods and Software
Vol. 20, Nos. 4–5, August–October 2005, 509–514

Taylor & Francis
Taylor & Francis Group

Interpolating greedy and reluctant algorithms

PIERLUIGI CONTUCCI*†, CRISTIAN GIARDINÀ†, CLAUDIO GIBERTI‡, FRANCESCO UNGUENDOLI§ and CECILIA VERNIA§

†Dipartimento di Matematica, Università di Bologna, Piazza di Porta S. Donato 5, 40127 Bologna, Italy
‡Dipartimento di Informatica e Comunicazione, Università dell'Insubria, via Mazzini 5, 21100 Varese, Italy
§Dipartimento di Matematica Pura ed Applicata, Università di Modena e Reggio Emilia, via Campi 213/B, 41100 Modena, Italy

(Received 2 October 2003; revised 20 April 2004; in final form 18 October 2004)

In a standard NP-complete optimization problem, we introduce an interpolating algorithm between the quick decrease along the steepest descent direction (greedy dynamics) and a slow decrease close to the level curves (reluctant dynamics). We find that, for a fixed elapsed computer time, the best performance of the optimization is reached at a special value of the interpolation parameter, considerably improving the results of the pure cases of greedy and reluctant.

Keywords: Combinatorial optimization; Statistical mechanics; Spin glass

Mathematical Subject Classification: 82B44; 82B80; 90C27

1. Introduction

Combinatorial optimization stands as one of the most fruitful fields in the intersection of applied and pure mathematics. It connects the theory of the computational hardness to the techniques widely used in the search of global minima for complex functionals, i.e. functions with many local minima. Over the years, there have been many strategies proposed to solve hard computational problems efficiently [1–3]. Among them the statistical mechanics approach [4,5] has opened new interesting perspectives. In this paper, we study the interpolation between two algorithms: the greedy algorithm and the reluctant algorithm (see ref. [6]). The first is the standard decrease along the deepest descent direction, while the second is the closest decrease to the level lines. In previous works [7,8], we have studied and compared the two algorithms, focusing on relaxation time and minimum reached level. We also observed how a simple convex interpolation between them could improve both their performances in large size regime. In this work, we push such analysis further by introducing a stochastic algorithm which smoothly interpolates between a greedy and a reluctant regime, depending on the value of a parameter λ: small λ plays the role of the greedy algorithm, while large λ that of the

*Corresponding author. Email: contucci@dm.unibo.it

Optimization Methods and Software
ISSN 1055-6788 print/ISSN 1029-4937 online
http://www.tandf.co.uk/journals
DOI: 10.1080/10556780500140177

reluctant. This allows us to obtain a better tuning between the two and especially a parameter optimization. The newly introduced algorithm is tested against a model which has become the standard of NP-complete problems [9,10]: the Sherrington–Kirkpatrick model of the mean field spin glass.[†] Our results confirm and extend those in refs. [7,8]. We find that the relaxation time grows linearly when the algorithm is close to the greedy regime and quadratically when it is close to the reluctant one, with a progressive condensation for large values of λ. The dynamics is then tested in the search of low energy configurations for fixed values of initial conditions where the reluctant dynamics works substantially better than any other. The main result of this work is then the minimization at fixed elapsed computer time. In this case, in fact, we find that in the small size regime the greedy component performs better than any other due to the short relaxation time and the fact that basically the dynamic is able to find the ground state or at least get very close to it. Moreover, and more interestingly, we find that increasing the system size does not lead to a uniform deterioration of the greedy performance toward an improvement of the reluctant one. We find, in fact, an optimal value of $\lambda \sim 10$, for which the lowest energy is reached against a rather poor performance of the greedy ($\lambda \sim 1$) and reluctant algorithm ($\lambda \sim 100$). This optimal value appears to be independent of the size.

2. The model and the algorithm

Let us consider the Sherrington–Kirkpatrick model [11] defined by the Hamiltonian

$$H(J,\sigma) = -\frac{1}{2}\sum_{i,j=1}^{N} J_{ij}\sigma_i\sigma_j, \tag{1}$$

where $\sigma_i = \pm 1$ for $i = 1, \ldots, N$ are spin variables which interact through an $N \times N$ symmetric matrix with J_{ij} independent, identically distributed Gaussian random couplings ($J_{ij} = J_{ji}$, $J_{ii} = 0$) with zero mean and variance $1/N$. We focus our attention on a stochastic energy-decreasing dynamics that, starting from any initial spin configuration at time $t = 0$ (which we choose at random with uniform distribution), ends up on a local energy minimum. The evolution rule is:

1. Let $\sigma(t) = (\sigma_1(t), \ldots, \sigma_N(t))$ be the spin configuration at time t.
2. Calculate the spectrum of energy change obtained by flipping the spin in position i, for $i = 1, \ldots, N$:
$$\Delta E_i = \sigma_i(t)\sum_{j\neq i} J_{ij}\sigma_j(t). \tag{2}$$
If $\Delta E_i > 0, \forall i = 1, \ldots, N$, then the algorithm stops ($\sigma(t)$ is a local minimum).
3. Generate a random number D with probability density
$$f(x) = \begin{cases} \lambda e^{\lambda x} & \text{if } x \leq 0 \\ 0 & \text{if } x > 0 \end{cases}, \quad \lambda > 0. \tag{3}$$

[†]See the 'P *vs.* NP Problem' description among the '*Millenium Problems*' at http://www.claymath.org/millennium/P_vs_NP/.

4. Select the site $i^\star$ associated with the closest energy change to the value D, i.e.:

$$i^\star = \left\{ i \in \{1, \ldots, N\} : \Delta E_{i^\star} = \min_{i \in \{1,\ldots,N\}} \{|\Delta E_i - D| : \Delta E_i < 0\} \right\}. \tag{4}$$

5. Flip the spin on site $i^\star$:

$$\sigma_i(t+1) = \begin{cases} -\sigma_i(t) & \text{if } i = i^\star \\ \sigma_i(t) & \text{if } i \neq i^\star. \end{cases} \tag{5}$$

The choice of the exponential distribution is standard in statistical mechanics and reflects the Gibbs–Boltzmann equilibrium ensemble. Moreover, in our problem it also turned out to be particularly convenient, thanks to the fact that it is tuned by a single parameter (λ). The considered algorithm generates a dynamics that, following a single spin flip decreasing energy trajectory, arrives at a single spin flip stable configuration, i.e. a configuration whose energy cannot be decreased by a single spin flip. The speed of convergence to local energy minima is tuned by λ, the control parameter in the probability distribution function for the move acceptance. Of course, the larger the λ, the bigger is the probability of doing small energy-decreasing steps, so that the trajectory will follow an evolution path close to level curves (reluctant). In contrast, small values of λ enrich the probability of large negative energy steps, which will quickly drive the dynamic to the end-point (greedy). In the next section, by varying the control parameter λ, we study the efficiency of the algorithm by measuring the average time to reach a metastable configuration and the lowest energy value found as the system size is increased.

3. Results

We performed a set of experiments for different values of N, starting from N initial conditions (for a system of size N) and averaging the data on $nreal = 1000$ disorder realizations. We probed basically two quantities to measure the performance of the algorithm:

- the average time (i.e. the number of spin flips) to reach a minimum energy level

$$\tau = \frac{1}{M} \sum_{i=1}^{M} t_i, \tag{6}$$

  with $M = N \cdot nreal$ and t_i, $i = 1, \ldots, M$ the time for each initial condition
- the lowest energy found (averaged over disorder)

$$H_N = \left\langle \frac{\min_\sigma H_N(J, \sigma)}{N} \right\rangle_{nreal}, \tag{7}$$

  where $\min_\sigma H_N(J, \sigma)$ is the minimum value of the energy of the metastable states attained starting from N initial conditions.

In figure 1, the relaxation time to a metastable configuration is analyzed. We represent τ as a function of N in the range [25, 300] for six distinct values of λ ($\lambda = 1, 10, 25, 45, 70, 100$), together with the best numerical fits. Because of high computational costs (which increase with λ), the case $N = 300$ is studied in detail only for $\lambda = 1$ and $\lambda = 10$. In contrast, the average time has good self-averaging properties so that, in order to have the trend for 'large' λ

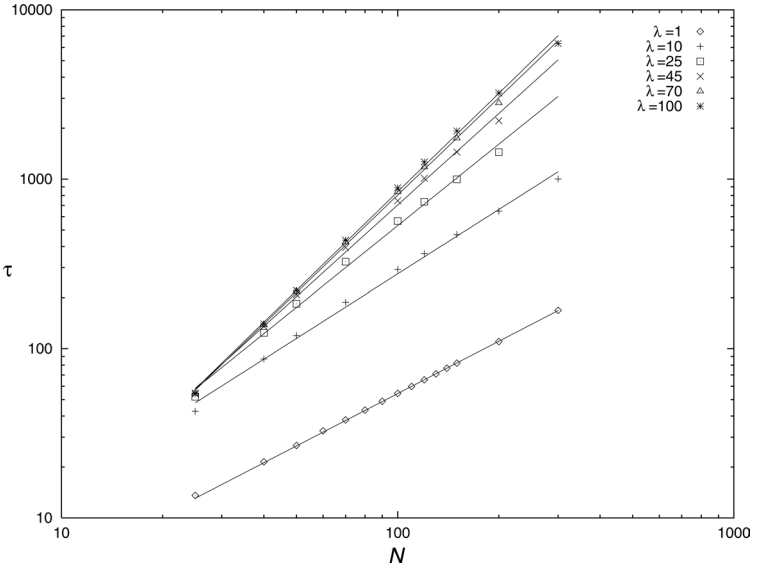

Figure 1. Average time τ to reach a metastable configuration as a function of N for $\lambda = 1$ ($\diamond$), $\lambda = 10$ ($+$), $\lambda = 25$ ($\square$), $\lambda = 45$ ($\times$), $\lambda = 70$ ($\triangle$), $\lambda = 100$ ($*$).

we focused on the case $N = 300$, $\lambda = 100$ with *nreal* $= 140$ disorder realizations instead of *nreal* $= 1000$. From figure 1 we observe the progressive increase of the slope in log–log scale from an almost linear law in N for $\lambda = 1$ ($\diamond$) to an almost quadratic one for $\lambda = 100$ ($*$). When comparing with the results of the previous analysis [7], we observe that this algorithm behaves as a 'smooth interpolation' between the two deterministic dynamics: 'greedy', that we obtain for $\lambda = 1$, and 'reluctant', here represented by $\lambda = 100$. The numerical fits of figure 1 are: $\tau_{\lambda=1}(N) \sim N^{1.027}$ ($\diamond$), $\tau_{\lambda=10}(N) \sim N^{1.263}$ ($+$), $\tau_{\lambda=25}(N) \sim N^{1.600}$ ($\square$), $\tau_{\lambda=45}(N) \sim N^{1.796}$ ($\times$), $\tau_{\lambda=70}(N) \sim N^{1.911}$ ($\triangle$), $\tau_{\lambda=100}(N) \sim N^{1.932}$ ($*$). The fits are quite good for all the cases except $\lambda = 10$ and $\lambda = 25$. In these cases, the quality of the fit is enhanced when excluding the data corresponding to $N = 25$ and $N = 40$. So we obtain $\tau_{\lambda=10}(N) \sim N^{1.184}$ and $\tau_{\lambda=25}(N) \sim N^{1.488}$.

Next, we measured the lowest energy found by the algorithm. In figure 2 we represent H_N as a function of N for different values of λ and for a fixed number of N initial conditions. While all the values of λ give basically the same results for small N, finding the ground state of the system, the best results for large N (>50) are obtained for $\lambda = 100$, which corresponds to deterministic reluctant dynamics. Therefore, this confirms that, for a fixed number of initial spin configurations, the algorithm that makes moves corresponding to the 'smallest' possible energy decrease is the most efficient in reaching low-energy states. In other words, the slower the better! However, this is reflected in an increasing cost for the computational time.

Indeed, when the analysis is focused on the performances for fixed elapsed time, the situation changes drastically. In figure 3, we compare the minimum energy values H_N, obtained by considering different system sizes and, for each of them, five different parameter values ($\lambda = 1, 10, 25, 45, 100$) for an elapsed time of 50 h of CPU on an IBM SP4. The system size $N = 350$ is studied only for $\lambda = 10, 25$ and 45. Each run (i.e. for fixed N and λ) consists of 1000 disorder realizations (J_{ij}), with the same CPU time length (3 min) assigned to each sample. The best result is obtained for the case $\lambda = 10$, which seems to be the best compromise to obtain a dynamical trajectory that is able to arrive deep enough with respect to energy levels, but without wasting all the time in the search for the slower of the possible paths. We note that this finding is in good agreement with the result of the previous analysis [7], where a convex

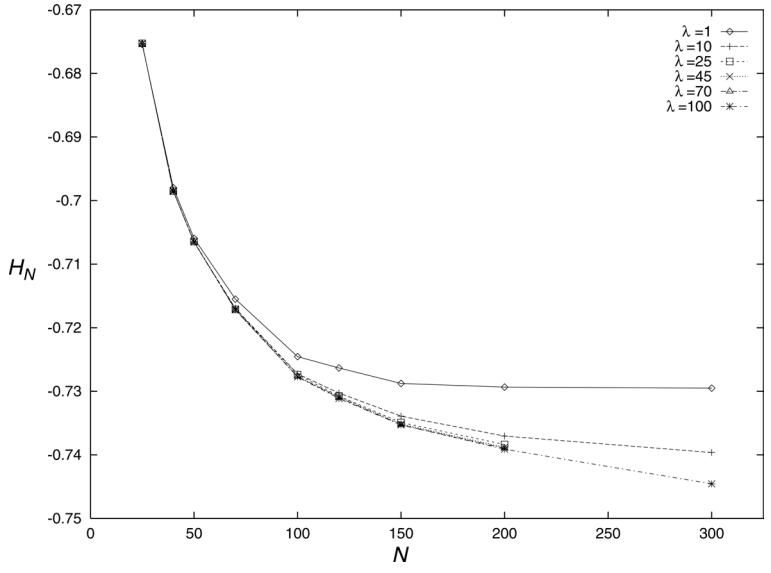

Figure 2. Lowest energy value H_N as a function of N obtained using a protocol with N initial conditions for 1000 disorder realizations for $\lambda = 1$ ($\diamond$), $\lambda = 10$ ($+$), $\lambda = 25$ ($\Box$), $\lambda = 45$ ($\times$), $\lambda = 70$ ($\triangle$), $\lambda = 100$ ($*$).

linear combination of reluctant (with probability P) and greedy (with probability $1 - P$) dynamics was considered. The optimal value $\lambda \sim 10$ is the one for which the relaxation time $\tau \sim N^{\alpha}$ grows with a scaling exponent α which is the closest, among the others, to the value $\alpha = 1.26$ of the optimal convex combination with $P = 0.1$.

We remark that figures 2 and 3 refer to different simulation protocols. While the first considers experiments with a fixed large number of initial conditions, the latter shows results for a fixed elapsed CPU time. In both cases convergence to metastable states is allowed.

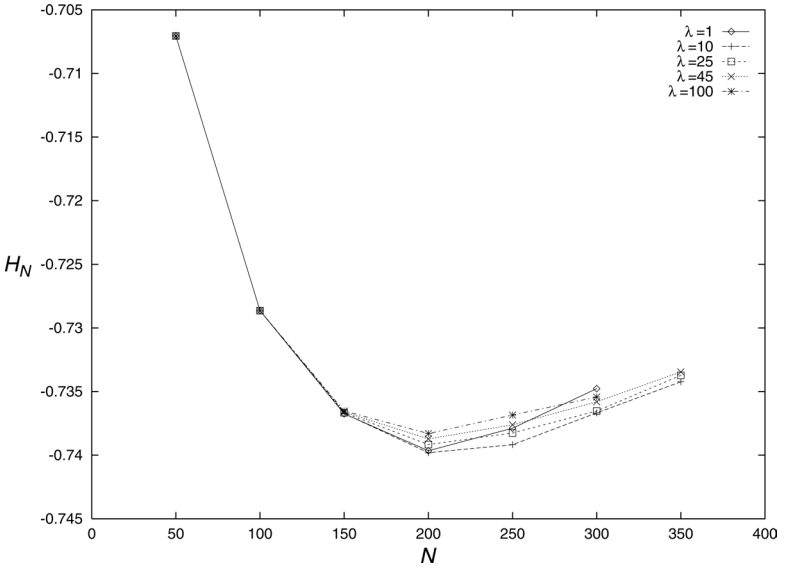

Figure 3. Lowest energy value H_N as a function of N for a fixed CPU time of 50 h on a IBM SP4 for $\lambda = 1$ ($\diamond$), $\lambda = 10$ ($+$), $\lambda = 25$ ($\Box$), $\lambda = 45$ ($\times$), $\lambda = 100$ ($*$).

Improvements of the greedy and reluctant algorithms is presently under study [12], by also permitting increase in energy with exponential decrease in time, in the very same spirit of the well-known simulated annealing strategies.

Acknowledgements

We thank the Cineca staff for technical support. We acknowledge Prof. S. Graffi for the encouragement and Prof. I. Galligani for inviting one of us (P.C.) to the Scientific Meeting 'Numerical Methods for Local and Global Optimization: Sequential and Parallel Algorithms', where this work was presented.

References

[1] Fletcher, R., 2000, *Practical Methods of Optimization* (2nd edn) (New York: John Wiley & Sons).
[2] Ahuja, R., Magnanti, T. and Orlin, J., 1993, *Network Flows* (New Jersey: Prentice Hall).
[3] Cook, W.J., Cunningham, W.H., Pulleyblank, W.R. and Schrijver, A., 1998, *Combinatorial Optimization* (New York: John Wiley & Sons).
[4] Mezard, M., Parisi, G., Virasoro, M.A., 1987, *Spin Glass Theory and Beyond* (Singapore: Word Scientific).
[5] Nishimori, H., 2001, *Statistical Physics of Spin Glasses and Information Processing* (New York: Oxford University Press).
[6] Parisi, G., 2003, On the statistical properties of the large time zero temperature dynamics of the SK model. *Fractals*, **11**(Suppl. 1), 161–171.
[7] Bussolari, L., Contucci, P., Degli Esposti, M. and Giardinà, C., 2003, Energy-decreasing dynamics in mean-field spin models. *Journal of Physics A (Mathematical and General)*, **36**, 2413–2421.
[8] Bussolari, L., Contucci, P., Giardinà, C., Giberti, C., Unguendoli, F. and Vernia, C., 2003, Optimization strategies in complex systems. Technical Report submitted to Science and Supercomputing at Cineca. Available online at: http://arxiv.org/abs/math.NA/0309058.
[9] Papadimitriou, C., 1994, *Computational Complexity* (Reading, Massachusetts: Addison-Wesley).
[10] Sipser, M., 1996, *Introduction to the Theory of Computation* (Boston: Brooks Cole).
[11] Sherrington, D. and Kirkpatrick, S., 1975, Solvable model of a spin glass. *Physics Review Letters*, **35**, 1792.
[12] Contucci, P., Giardinà, C., Giberti, C. and Vernia, C., Annealed greedy and reluctant dynamics, to be published in *Mathematical Models and Methods in Applied Sciences*.