Using Hysteresis for Optimization

G. Zaránd,[1,2] F. Pázmándi,[2,3,5] K. F. Pál,[4] and G. T. Zimányi[3]

[1]*Lyman Physics Laboratory, Harvard University, Cambridge, Massachusetts 02145*
[2]*Research Group of the Hungarian Academy of Sciences, Institute of Physics, TU Budapest, H-1521 Hungary*
[3]*Department of Physics, University of California, Davis, California 95616*
[4]*Institute of Nuclear Research of the Hungarian Academy of Science, P.O. Box 51, Debrecen, H-4001 Hungary*
[5]*Department of Theoretical Physics, University of Debrecen, P.O. Box 5, Debrecen, H-4010 Hungary*

(Received 8 May 2001; published 23 September 2002)

We propose a new optimization method based on a demagnetization procedure well known in magnetism. We show how this procedure can be applied as a general tool to search for optimal solutions in any system where the configuration space is endowed with a suitable "distance." We test the new algorithm on frustrated magnetic models and the traveling salesman problem. We find that the new method successfully competes with similar basic algorithms such as simulated annealing.

DOI: 10.1103/PhysRevLett.89.150201 PACS numbers: 02.60.Pn, 75.10.Nr, 75.50.Lk

It was observed a long time ago that disordered materials can be brought into a remarkably stable state through annealing, i.e., cooling down the material rather slowly. This simple observation inspired Kirkpatrick, Gelatt, and Vecchi in their pioneering work [1] to investigate how close this annealing procedure takes models with glassy properties to their ground state, and it led them to the invention of the by now widely used simulated annealing (SA) procedure. The SA revealed the crucial role the external noise can play in optimization: Thermal noise can help to escape high-energy local minima. In the present work we investigate another procedure that is commonly used to demagnetize disordered magnets and is experimentally known to result in a very stable state: the application of an oscillating external field (see Fig. 1). This procedure makes use of another type of noise which is typical in magnetic systems, namely, *random external fields*. As we demonstrate below for various models, a simple generalization of this zero-temperature "ac demagnetization" (ACD) is able to give systematically better and better approximations to the ground state of these models and is in many cases 5–10 times faster than SA. We show how this method can be applied to practically *any* disordered model, thereby resulting in a new optimization procedure, which we call *hysteretic optimization* (HO).

Finding optimal solutions of complex problems depending on a large number of parameters is an equally important and difficult task [2]. Examples range from integrated circuit design, through portfolio selection on the stock market [3] and calculating protein folding, to teaching artificial neural networks, to name a few. The simultaneous presence of randomness and frustration is what makes these problems so hard: Disorder is caused by the nonregular dependence of the quality of the solution on the configuration, and frustration is brought in by the competition of mutually exclusive different "good" properties. As a result, on one hand, a naive search often gets stuck in spurious minima while, on the other hand, comparably good solutions can be found with quite different configurations.

It is important to note that most hysteretic systems fulfill the above requirements of complexity. Hysteresis implies the presence of many metastable states caused mainly by disorder. In the case of magnetic materials, the other important ingredient, frustration, is furnished by the magnetostatic interaction, which can be ferromagnetic or antiferromagnetic depending on the relative orientation of the dipoles. This analogy between magnetic systems and optimization suggests that part of the knowledge accumulated through the decades by hysteresis research [4,5] will eventually prove useful in optimization (and vice versa). Indeed, a simple but very frequently used hysteresis model, the Preisach model [4–6], can be put in its ground state by ACD.

In this Letter, we proceed by studying in detail the energetics of ACD for spin glass models. Then we show how this simple method can be modified to obtain a hysteretic optimization technique, which systematically approaches the ground state [7], and we compare the new algorithm with SA. We also present how HO can be combined with cluster renormalization and apply the general formalism of HO for the traveling salesman problem (TSP).

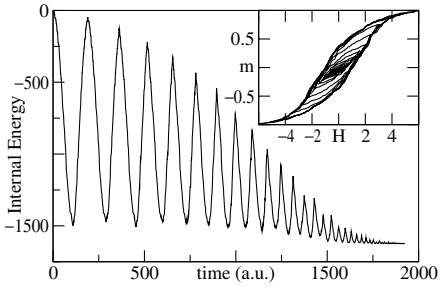

FIG. 1. The total internal energy (without the Zeeman term) of a three-dimensional ($L = 10$) Edwards-Anderson spin glass during ACD. The inset shows the corresponding magnetization curve.

Many optimization problems can be formulated in terms of interacting Ising spins ($\sigma = \pm 1$). Therefore, we studied first two classical Ising spin glass models [8] with a Hamiltonian:

$$\mathcal{H} = -\frac{1}{2}\sum_{i,j}^{N} J_{ij}\sigma_i\sigma_j - H\sum_{i}^{N}\xi_i\sigma_i. \qquad (1)$$

In the case of the Sherrington-Kirkpatrick (SK) model for any (i, j) pair $J_{ij} = z_{ij}/\sqrt{N}$, where z_{ij} is a random Gaussian number with zero mean and unit variance. In the other spin glass we considered, the three-dimensional Edwards-Anderson (EA) model, $J_{ij} = z_{ij}$, but only for the nearest-neighbor couplings. The direction $\xi_i = \pm 1$ of the external field H randomly changes from site to site. Because of a spin-gauge symmetry, any choice of ξ_i is equivalent, as long as it is not correlated with J_{ij}. Using this symmetry, and also the analogy of the magnetic energy from Eq. (1), we call $m = 1/N\sum \xi_i\sigma_i$ magnetization throughout the Letter. These systems are the focus of research both in magnetism and in optimization. In magnetism their dynamics and the nature of the ordered state is still highly debated [9,10]. From the optimization point of view, they are interesting because they are considered as "hard" problems [11]. The major difference between the two models lies in the range of interactions. The short-ranged EA model may show some kind of clusterization ("droplets"), which is very unlikely for the infinite-ranged SK model.

We tested first the energetic properties of ACD on the SK model [12]. Many properties of this model can be calculated exactly [8], and it is an ideal test system. We know, e.g., that its ground state energy is -0.765/spin as $N \to \infty$. Starting from a random initial state, one can simply *quench* the system in the following way: (a) Identify all unstable spins opposite the local fields $h_i = \sum_j J_{ij}\sigma_j + H\xi_i$, (b) flip randomly one of them and update the local fields, (c) repeat (a) and (b) until all spins are stable. Such a quench ends up with (locally) stable states of energy around -0.70 per spin at $H = 0$.

To perform the ACD we start with an external field $H = H_1 \equiv \max\{-\xi_i\sum_j J_{ij}\xi_j\}$ that saturates the spins (i.e., the $\sigma_i = \xi_i$ state with $m = 1$ is stable). Then we decrease the field H until a spin becomes unstable. We start then an avalanche by flipping this spin and "quench" the system to a nearby stable state [13]. We continue this procedure until H reaches the "turning point" $H_2 = -\gamma_1 H_1$; then we start to *increase* the field with the same procedure until we reach $H_3 = -\gamma_2 H_2$. We turn back again and again at points $H_n = -\gamma_{n-1}H_{n-1}$, until the amplitude of the loop, $|H_n|$, becomes smaller than the field necessary to flip a spin. Then we set $H = 0$. The resulting "spiraling" magnetization curve is shown in the inset of Fig. 1. We found that introducing small fluctuations in the γ_i's increased the efficiency. All results presented here were obtained with γ_i's generated randomly between 0.8 and 1 with a uniform distribution, with an average $\gamma \equiv \langle\gamma_i\rangle = 0.9$. We found that, as expected, lower γ gives poorer results, but a γ even closer to 1 did not make a considerable improvement.

Figure 2 shows for different system sizes the probability of finding a configuration close to the ground state of the SK model with ACD [14]. For each N we generated several realizations of the couplings and ran ACD many times with different random field directions $\{\xi_i\}$. Energies were measured from the exact ground state [14] of a given realization of couplings (instance). The distributions presented are averaged over instances and field directions $\{\xi_i\}$. The finite intercept at $E = 0$ for smaller systems signals a finite probability of finding the ground state, but this probability goes to zero as the size increases: for very large systems ACD typically gives a state with an energy $0.012 \times N$ above the ground state [14].

The probability density $P(E)$ of finding a state with energy E above the ground state produced by ACD for the $N = 1000$ SK model is shown by the symbols in Fig. 3. These measurements were performed for a single sample (instance). The ACD gives remarkable improvement with respect to the quench, but in both cases the final state is definitely above the ground state. However, ACD runs 3–10 times faster than SA (see Figs. 3 and 4), and can be a viable alternative in applications where a large number of optimizations is required, each on relatively smaller systems.

As mentioned earlier, unlike SA, increasing the factor γ closer to 1, we find only a limited effect on the quality of solutions obtained by ACD. This is a serious problem, and finding the cause of it is an important subject for future research. We speculate, however, that the answer is hidden in the random, but "frozen" correlations between our noise term ξ_i and the generated low-lying states. These correlations may prevent us from reaching certain parts of the phase space. If so, all we have to do is to introduce different ξ_i-s, but without destroying the good correlations of previous ξ_i-s.

Maybe the simplest way to achieve this is "shaking up" the system: After finishing an ACD, H being zero, we are free to change ξ_i to a different noise "direction" ξ_i', while the final spin configuration σ_i^0 remains stable. Starting from this state, we increase the field in small

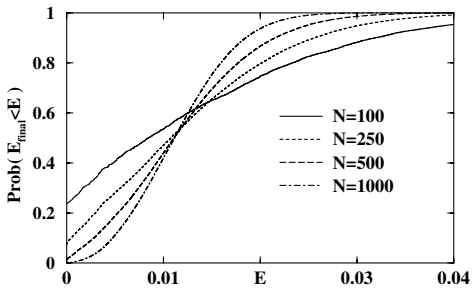

FIG. 2. Size dependence of the integrated energy distribution of states produced by ACD for the SK model. E is the energy per spin measured from the ground state of each sample.

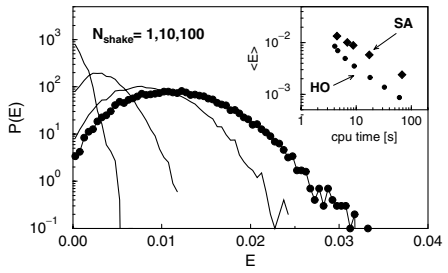

FIG. 3. Probability density of finding a state of an SK model ($N = 1000$, $H_{\text{shake}} = 0.9$) with energy E per spin above the ground state by ACD (symbols) and HO with 1, 10, and 100 shakeups (solid lines). Inset: Average energy per spin measured from the ground state vs running time for the SK model ($N = 1000$).

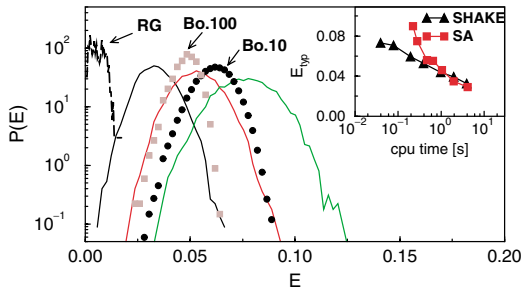

FIG. 4 (color online). Probability density of finding a state with energy E per spin above the ground state for the three-dimensional ($L = 10$) Edwards-Anderson model with HO with $n = 1$, 10, and 100 consecutive shakeups (solid lines) and $H_{\text{shake}} = 2.2$. Symbols: Best of $n = 10$, and 100 single shakeups. Dashed lines: HO combined with renormalization group (RG) techniques. Inset: Typical energies as a function of running times for SA and HO with shakeups.

steps to H_{shake}, quenching the spins after each step as before. From this point we start a new ACD by decreasing the field to $-\gamma_1 H_{\text{shake}}$, and turning back, etc. At the end of the shakeup we get a new state σ_i^1, and H is zero again, so we can choose a new ξ_i'', and start another shakeup. While subsequent shakeups with appropriate H_{shake} improve the solution on average, it is obviously helpful to get rid of "bad fluctuations"; i.e., in case σ_i^{n+1} had higher energy than σ_i^n, we would start the new shakeup from the latter again. It is this combination of ACD and shakeups that we call *hysteretic optimization* [15].

The exact choice of the shaking amplitude H_{shake} is best obtained by testing it with a few discrete values: too small loops make hardly any change and too large field values destroy previous good correlations and take us back to ACD results. Our practice shows that the best choice for H_{shake} is a field somewhat larger than the coercive field.

Figures 3 and 4 summarize our results of the HO calculations on the SK and EA models. Already the first few shakeups give substantial improvement over ACD. Increasing further the number of shakeups is less and less effective, but we get systematically closer and closer to the ground state. Figure 4 also shows that results for the "best of n shakeups" are inferior to "n consecutive shakeups," for the latter build up gradually the "good correlations," as explained earlier. We emphasize that applying shakeups after an ACD is straightforward, requiring very little extra coding and computer time. A systematic comparison of ACD, shakeup HO, and SA is shown in Figs. 3 and 4.

In the case of the short-range EA model (inset of Fig. 4) shakeups seem less effective, and for longer running times SA reaches the efficiency of HO. However, typical optimization problems are often closer to the long-range SK model where no particular dimensionality is present (e.g., stock market) [3], for which HO appears to be better than SA even for long running times (Fig. 3).

We have to emphasize that, as well as SA, the HO presented in this Letter should be considered as a *building block,* and it can be efficiently combined with other optimization methods, such as genetic algorithms [17], cluster RG techniques [18], or both [19]. To demonstrate this, we combined HO with the RG in the following way: We generated clusters from $n(\sim 10$–$20)$ different low-energy configurations obtained by HO and using the clusterization technique of Ref. [19]. Using cluster spins simplifies the problem not just by decreasing the number of degrees of freedom, but also by removing dynamical barriers: turning over a bigger cluster spin by spin might be energetically expensive because of the strong couplings inside the cluster. For the effective system of cluster spins HO can be applied again to generate n new (cluster) configurations and to define clusters of clusters, and so on. We used this relatively large n [20] on each RG level, and consequently many RG steps to avoid the necessity of including a genetic algorithm.

It is worth noting that there is a freedom in introducing the external field for the clusters, since the "magnetic moment" μ of a cluster can be arbitrarily chosen. The simplest choice is $\mu = 1$ for all clusters, but it might be useful to scale the local moments to ensure that small and large clusters can compete with each other. The success of the RG approach for the EA model is mostly due to the short-range nature of the interactions (see Fig. 4). We find that this HO + RG method is able to reach the ground state in a time comparable to a recent very efficient genetic algorithm developed specifically for this problem [9,16].

Now we show how to apply the present algorithm to practically any optimization problem. In an optimization problem one has to minimize some cost function $W(P)$, which is a mapping from some configuration space $\{P\}$ to real numbers. In order to apply HO to a system we need three important ingredients: (i) *dynamics.* This is already sufficient to do SA; however, for the HO we need also a (ii) *distance* $d(P, Q)$ over the configuration space, and (iii) two *reference states* $R_{\pm}$. These latter will be used to induce a random external field.

We have a freedom in choosing these ingredients. However, a good dynamics must be such that an elementary

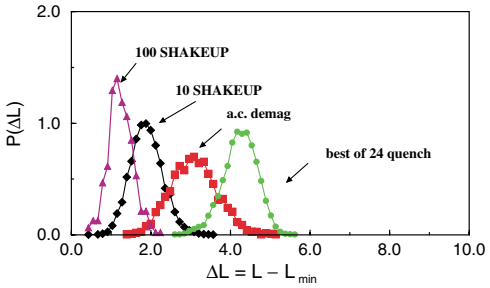

FIG. 5 (color online). Distribution of length generated by HO for a two-dimensional TSP with $N = 100$. For comparison, we also show the best of 24 quenches, taking the same time as ACD.

step connects configurations with approximately equal cost functions, and its successive applications connect all configurations. Similarly, a good distance corresponding to a given dynamics is such that distances do not change much during a single step.

Having defined all these quantities we define the HO as the demagnetization of the Hamiltonian

$$\mathcal{H}(P) = W(P) + \sum_{\alpha=\pm} \alpha H \Theta(\alpha H) d(P, R_\alpha), \qquad (2)$$

with $\Theta(x)$ being the step function. For the spin problems considered here the configurations are $P = \{\sigma_i\}$, the dynamics is single spin flip, the distance is $d(\{\sigma_i\}, \{\tilde{\sigma}_i\}) = \sum_i |\sigma_i - \tilde{\sigma}_i|$, and the two reference states are $R_\pm = \{\pm\xi_1, \ldots, \pm\xi_N\}$. With these definitions Eq. (2) reduces to Eq. (1), and the general HO becomes the one we applied to the spin problems. The basic role of the external field in Eq. (2) is to force the system close to the states $R_\pm$ as $H \to \pm\infty$.

We now show how to apply this method for the TSP. In this classical optimization problem the aim is to find the shortest path that goes through each of a given set of $i = 1, \ldots, N$ cities located at positions $\mathbf{x}_i$, visiting just once each of them. The configuration space is given by all possible permutations P of the cities, and the cost function is simply the total length of the path, $W(P) = \sum_{i=1}^{N} |\mathbf{x}_{P(i)} - \mathbf{x}_{P(i+1)}|$. As an elementary step we used the interchange of two cities [21], and a distance

$$d(P, Q) \equiv \sum_{i}^{N} |\boldsymbol{\Delta}_i(P) - \boldsymbol{\Delta}_i(Q)|, \qquad (3)$$

where $\boldsymbol{\Delta}_i(P) = \mathbf{x}_{P(i+1)} - \mathbf{x}_{P(i)}$ denotes the vector connecting the cities $P(i)$ and $P(i + 1)$. Our results for the traveling salesman are summarized in Fig. 5. For the reference states $P_\pm$ we used two randomly generated paths. We find a systematic improvement as the number of shakeups increases, and we get better and better results in this case, too.

SA has the nice property of self-consistently telling us how the annealing schedule should be set: the specific heat is proportional to the fluctuations, so one might want to spend more time in the temperature regions of high specific heat. A similar measure of accuracy for HO might be the *measured* Preisach function. This function can be obtained numerically from first order reversal curves [5], and once at hand, it can guide the annealing of the loops. We believe that this function has a lot of information about the metastable states of the system, although future research is needed to identify these.

This research has been supported by NSF Grants No. DMR99-81283, No. DMR97-14725, and No. DMR99-85978 and Hungarian Grants No. OTKA F030041, No. T038162, and No. T037212. G. Z. is an Eötvös fellow.

[1] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, Science **220**, 671 (1983).
[2] C. H. Papadimitriou and K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity* (Prentice-Hall, Englewood Cliffs, NJ, 1982).
[3] S. Gallucio, J.-P. Bouchaud, and M. Potters, Physica (Amsterdam) **259A**, 449 (1998); A. Gábor and I. Kondor, Physica (Amsterdam) **274A**, 222 (1999).
[4] G. Bertotti, *Hysteresis and Magnetism* (Academic Press, New York, 1998).
[5] E. Della Torre, *Magnetic Hysteresis* (IEEE Press, Piscataway, NJ, 1999).
[6] F. Preisach, Z. Phys. **94**, 277 (1935).
[7] We have no proof of *convergence* to the ground state.
[8] M. Mézard, G. Parisi, and M. A. Virasoro, *Spin-Glass Theory and Beyond* (World Scientific, Singapore, 1987).
[9] M. Palassini and A. P. Young, Phys. Rev. Lett. **85**, 3017 (2000); J.-F. Bouchaud (unpublished).
[10] A. A. Middleton, Phys. Rev. B **63**, 060202(R) (2001).
[11] A. K. Hartmann, Physica (Amsterdam) **224A**, 480 (1996).
[12] G. Bertotti and M. Pasquale, J. Appl. Phys. **67**, 5255 (1990).
[13] Changing H in *finite steps* could produce a substantial increase in the speed of the algorithm for systems with small avalanches like the EA model.
[14] We used the method of [16] to obtain the ground state for all models.
[15] For a detailed algorithmic description, see G. Zarand *et al.*, cond-mat/0109359.
[16] K. F. Pal, Physica (Amsterdam) **223A**, 283 (1996); **233A**, 60 (1996).
[17] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning* (Addison-Wesley, Reading, MA, 1989).
[18] N. Kawashima and M. Suzuki, J. Phys. A **25**, 1055 (1992).
[19] J. Houdayer and O. C. Martin, Phys. Rev. Lett. **83**, 1030 (1999), used RG combined with a genetic algorithm.
[20] This cluster RG can be solved exactly for the one-dimensional EA spin glass.
[21] While more complicated dynamical steps [1] can be more efficient, here we chose a less efficient dynamics which allows one to more clearly judge the performance of HO. The ground state of the TSP was found using twist dynamics [1].