Optimization with Extremal Dynamics

STEFAN BOETTCHER
Physics Department, Emory University, Atlanta, Georgia 30322; e-mail: sboettc@emory.edu

ALLON G. PERCUS
Computer and Computational Sciences Division, Los Alamos National Laboratory, Los Alamos, NM 87545; e-mail: percus@lanl.gov

Received August 20, 2002; revised January 15, 2003; accepted January 15, 2003

A local-search heuristic for finding high-quality solutions for many hard optimization problems is explored. The method is inspired by recent progress in understanding far-from-equilibrium phenomena in terms of self-organized criticality, *a concept introduced to describe emergent complexity in physical systems. This method, called* extremal optimization, *successively replaces the value of extremely undesirable variables in a sub-optimal solution with new, random ones. Large, avalanche-like fluctuations in the cost function emerge dynamically. These enable the search to effectively scaling barriers to explore local optima in distant neighborhoods of the configuration space while eliminating the need to tune parameters. Drawing upon models used to simulate the dynamics of granular media, evolution, or geology, extremal optimization complements approximation methods inspired by equilibrium statistical physics, such as* simulated annealing. *This method is very general and so far has proved competitive with—and even superior to—more elaborate general-purpose heuristics on testbeds of constrained optimization problems with up to 10^5 variables, such as bipartitioning, coloring, and spin glasses. Analysis of a model problem predicts the only free parameter of the method in accordance with all experimental results. © 2003 Wiley Periodicals, Inc.**

Key Words: extremal optimization; criticality; simulated annealing; punctuated equilibrium

Many natural systems have, without any centralized organizing facility, developed into complex structures that optimize their use of resources in sophisticated ways [1]. Biological evolution has formed efficient and strongly interdependent networks in which resources rarely go to waste. Even the morphology of inanimate landscapes exhibits patterns that seem to serve a purpose, such as the efficient drainage of water [2,3].

Natural systems that exhibit self-organizing qualities often possess a common feature: a large number of strongly coupled entities with similar properties. Hence, at some coarse level they permit a statistical description. An external resource (sunlight in the case of evolution) drives the system which then takes its direction purely by chance. Like

descending water breaking through the weakest of all barriers in its wake, biological species are coupled in a global comparative process that persistently washes away the least fit. In this process, unlikely but highly adapted structures surface inadvertently. Optimal adaptation thus emerges naturally, from the dynamics, simply through a selection *against* the extremely "bad." In fact, this process prevents the inflexibility inevitable in a controlled breeding of the "good."

Various models relying on extremal processes have been proposed to explain the phenomenon of self-organization [4]. In particular, the Bak-Sneppen model of biological evolution is based on this principle [5,6]. Assuming an unspecified interdependency between species, it produces salient nontrivial features of paleontological data such as broadly distributed lifetimes of species, large extinction events, and punctuated equilibrium.

In the Bak-Sneppen model, species are located on the sites of a lattice, and have an associated "fitness" value between 0 and 1. At each time step, the one species with the smallest value (poorest degree of adaptation) is selected for a random update, having its fitness replaced by a new value drawn randomly from a flat distribution on the interval [0, 1]. But the change in fitness of one species impacts the fitness of interrelated species. Therefore, all of the species at neighboring lattice sites have their fitness replaced with new random numbers as well. After a sufficient number of steps, the system reaches a highly correlated state known as self-organized criticality (SOC) [7]. In that state, almost all species have reached a fitness above a certain threshold. These species, however, possess *punctuated equilibrium:* only one's weakened neighbor can undermine one's own fitness. This coevolutionary activity gives rise to chain reactions called "avalanches," large fluctuations that rearrange major parts of the system, potentially making any configuration accessible.

Although coevolution may not have optimization as its exclusive goal, it serves as a powerful paradigm. We have used it as motivation for a new approach to approximate hard optimization problems [8]. The heuristic we have introduced, called extremal optimization (EO), follows the spirit of the Bak-Sneppen model, updating those variables that have among the "worst" values in a solution and replacing them by random values without ever explicitly improving them.

To introduce EO, let us consider a spin glass [9] as a specific example of a hard optimization problem. It consists of a d-dimensional hyper-cubic lattice of length L with periodic boundary conditions, with a spin variable $x_i \in \{-1, 1\}$ at each site i, $1 \leq i \leq n\ (=L^d)$. A spin is connected to each of its nearest neighbors j via a bond variable $J_{ij} \in \{-1, 1\}$, assigned at random. The configuration space Ω consist of all configurations $S = (x_1, \ldots, x_n) \in \Omega$, where $|\Omega| = 2^n$.

We wish to minimize the cost function, or Hamiltonian

$$C(S) = H(x) = -\frac{1}{2}\sum_{\langle i,j\rangle} J_{ij}x_i x_j, \tag{1}$$

where the sum extends over all nearest-neighbor pairs of spins. Due to frustration [9], ground state configurations $S_{\min}$ are hard to find, and it has been shown that for $d > 2$ the problem is among the hardest optimization problems [10].

To find near-optimal solutions for a particular optimization problem, EO performs a neighborhood search on a single configuration $S \in \Omega$. As in the spin problem in Eq. (1), S consists of a large number n of variables x_i. We assume that we can define for each S a neighborhood $N(S)$ that rearranges the state of merely a small number of the variables. Those neighborhoods are a characteristic of a local search, in contrast to a genetic algorithm, say, where crossovers may effect $O(n)$ variables on each update. The cost $C(S)$ is assumed to consist of the individual cost contributions, or "fitnesses," λ_i for each variable x_i (analogous to the fitness values in the Bak-Sneppen model [5,8]). The fitness of each variable assesses its contribution to the total cost:

$$C(S) = -\sum_{i=1}^{n} \lambda_i. \tag{2}$$

Typically, the fitness λ_i depends on the state of x_i in relation to connected variables.

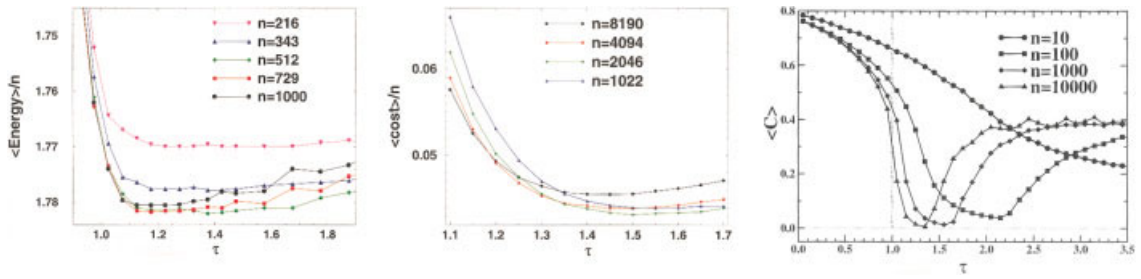
For example, for the Hamiltonian in Eq. (1), we assign to each spin x_i the fitness

$$\lambda_i = x_i\left(\frac{1}{2}\sum_j J_{ij}x_j\right), \tag{3}$$

so that Eq. (2) is satisfied. Each spin's fitness thus corresponds to (the negative of) its local energy contribution to the overall energy of the system. In similarity to the Bak-Sneppen model, EO then proceeds through a neighborhood search of Ω by sequentially changing variables with "bad" fitness on each update, for instance, via single spin-flips. After each update, the fitnesses of the changed variable and of all its connected neighbors are reevaluated according to Eq. (3). The basic EO algorithm proceeds as follows:

1. Initialize configuration S at will; set $S_{\text{best}} := S$.
2. For the "current" configuration S,
   (a) evaluate λ_i for each variable x_i,
   (b) find j satisfying $\lambda_j \leq \lambda_i$ for all i, i.e., x_j creates the "worst fitness,"
   (c) choose $S' \in N(S)$ such that j *must* change its state,
   (d) accept $S := S'$ *unconditionally,*
   (e) if $C(S) < C(S_{\text{best}})$ then set $S_{\text{best}} := S$.

FIGURE 1



Plot of the average costs obtained by EO for a $\pm J$ spin glass (left), for graph bipartitioning (center), and for the solvable model from [14], all as a function of τ. In the solvable model, we merely had to average runs for each n over initial conditions. For the real problems a number of instances were generated at each n. For each instance, 10 different EO runs were performed at each τ. The results were averaged over runs and over instances. Although both problems are quite distinct, in either case the best results are obtained at a value of τ that behaves according to Eq. (6), as predicted by the model.

3. Repeat at step (2) as long as desired.
4. Return S_{best} and $C(S_{\text{best}})$.

The algorithm operates on a single configuration S at each step. Each variable x_i in S has a fitness, of which the "worst" is identified. This *ranking* of the variables provides the only measure of quality on S, implying that all other variables are "better" in the current S. In the move to a neighboring configuration, typically only a small number of variables change state, so only a few connected variables need to be re-evaluated [step (2a)] and re-ranked [step (2b)]. Note that there is not a single parameter to adjust for the selection of better solutions. It is merely the memory encapsulated in the ranking that directs EO into the neighborhood of increasingly better solutions. Similar to the Bak-Sneppen model, those "better" variables possess punctuated equilibrium: their memory only get erased when they happen to be connected to one of the variables forced to change. On the other hand, in the choice of move to S', there is no consideration given to the outcome of such a move, and not even the worst variable x_j itself is guaranteed to improve its fitness. Accordingly, large fluctuations in the cost can accumulate in a sequence of updates. Merely the bias *against* extremely "bad" fitnesses produces improved solutions.

Tests have shown that this basic algorithm is very competitive for optimization problems where EO can choose randomly among many $S' \in N(S)$ that satisfy step (2c), as is the situation for graph bipartitioning [8,11]. But in cases such as the single spin-flip neighborhood for the spin Hamiltonian, focusing on only the worst fitness [step (2b)] leads to a deterministic process, leaving no choice in step (2c): If the "worst" spin x_j *has* to flip and any neighbor S' differs by only one flipped spin from S, it must be $S' = (S/x_j) \cup \{-x_j\}$. This deterministic process inevitably will get stuck near some poor local minimum. To avoid these "dead ends" and to improve results [8], we introduce a single parameter into the algorithm. To this end, we rank all x_i according to fitness λ_i, i.e., we find a permutation Π of the variable labels i with

$$\lambda_{\Pi(1)} \leq \lambda_{\Pi(2)} \leq \cdots \leq \lambda_{\Pi(n)}. \tag{4}$$

The worst variable x_j [step (2b)] is of rank 1, $j = \Pi(1)$, and the best variable is of rank n. Now, consider a scale-free probability distribution over the *ranks* k,

$$P_k \propto k^{-\tau}, \quad 1 \leq k \leq n, \tag{5}$$

for a fixed value of the parameter τ. At each update, select a rank k according to P_k. Then, modify step (2c) so that the variable x_j with $j = \Pi(k)$ changes its state.

For $\tau = 0$, this "τ-EO" algorithm is simply a random walk through Ω. Conversely, for $\tau \to \infty$, the process approaches a deterministic local search, only updating the lowest-ranked variable, and is bound to reach a dead end (see Figure 1). However, for finite values of τ the choice of a *scale-free* distribution for P_k in Eq. (5) ensures that no rank gets excluded from further evolution, while still maintaining a bias against variables with bad fitness. In all problems studied so far, a value of

$$\tau - 1 \sim 1/\ln n \quad (n \to \infty) \tag{6}$$

seems to work best [12,13]. We have studied a simple model problem for which the asymptotic behavior of τ-EO can be solved exactly [14]. The model reproduces Eq. (6) exactly in cases where the model develops a "jam" amongst its variables, which is quite a generic feature of frustrated systems: After many update steps most variables freeze into a near-perfect local arrangement and resist further change, while a finite fraction remains frustrated in a poor local arrangement. More and more of the frozen (slow) variables have to be dislocated collectively to accommodate the frustrated

TABLE 1

Best Cutsizes (and Allowed Runtime) for a Testbed of Large Graphs

Large Graph	GA	τ-EO	Ref. [16]	p-METIS
Hammond ($n = 4720$; $c = 5.8$)	90 (1s)	90 (42s)	97 (8s)	92 (0s)
Barth5 ($n = 15606$; $c = 5.8$)	139 (44s)	139 (64s)	146 (28s)	151 (0.5s)
Brack2 ($n = 62632$; $c = 11.7$)	731 (255s)	731 (12s)	—	758 (4s)
Ocean ($n = 143437$; $c = 5.7$)	464 (1200s)	464 (200s)	499 (38s)	478 (6s)

GA results are the best reported [15] (at 300 MHz). τ-EO results are from our runs (at 200 MHz). Comparison data for three of the large graphs are due to results from heuristics in Ref. [16] (at 50 MHz). METIS is a partitioning program based on hierarchical decomposition instead of local search [17], obtaining extremely fast deterministic results (at 200 MHz).

(fast) variables before the system as a whole can improve its state. In this highly correlated state, slow variables block the progression of fast variables, and a "jam" emerges. And our asymptotic analysis of the flow equations for a jammed system indeed reproduces Eq. (6).

GRAPH BIPARTITIONING PROBLEM

In the graph bipartitioning problem (GBP), variables x_i are given by a set of n vertices, where n is even. "Edges" connect certain pairs of vertices to form an instance of a graph. The problem is to find a way of partitioning the vertices into two subsets, each constrained to be *exactly* of size $n/2$, such that a *minimal* number of edges cut across the partition. The cost function $C(S)$ (called "cutsize") counts the number of such edges. Instances are typically parameterized by the average connectivity c of its vertices.

To implement τ-EO for the GBP, we attribute to each vertex i a local cost $\lambda_i = -b_i/2$, where b_i is the number of "bad" edges crossing the partition (equally shared with the vertex on the other end of that edge). Note that Eq. (2) is satisfied. The simplest neighborhood $N(S)$ for the GBP is an "exchange" of one vertex from each subset. For EO we choose two numbers k_1, k_2 according to Eq. (5) and exchange x_{j_1} and x_{j_2} with $j_1 = \Pi(k_1)$ and $j_2 = \Pi(k_2)$ as in Eq. (4), subject to the restriction that x_{j_1} and x_{j_2} be in opposite subsets. For the sizes of graphs studied, a value of $\tau = 1.4$–1.6 worked best [see Figure 1, middle]. Table 1 summarizes τ-EO's results on large-n graphs, using $\tau = 1.4$ and best-of-10 runs.

Studies on the average rate of convergence toward better-cost configurations as a function of runtime t indicate power-law convergence, roughly like $C(S_{\text{best}})_t \sim C(S_{\min}) + At^{-0.4}$ [13], also found by Ref. [18]. Of course, it is not easy to assert for graphs of large n that those runs in fact converge closely to the optimum $C(S_{\min})$, but finite-size scaling analysis for random graphs seems to justify that expectation [13].

In an extensive numerical study on random and geometric graphs [11] we have shown that τ-EO outperforms simulated annealing (SA) [19,20] significantly near phase transitions [21], where cutsizes first become nonzero. To this end, we have compared the averaged best results obtained for both methods for a large number of instances for increasing n at a fixed parameter setting ($\tau = 1.4$ for EO). Figure 2 shows that the SA implementation produces increasingly worse results near the phase transition relative to EO. In turn, it was found that EO reproduces many features of the transition quite accurately [11].

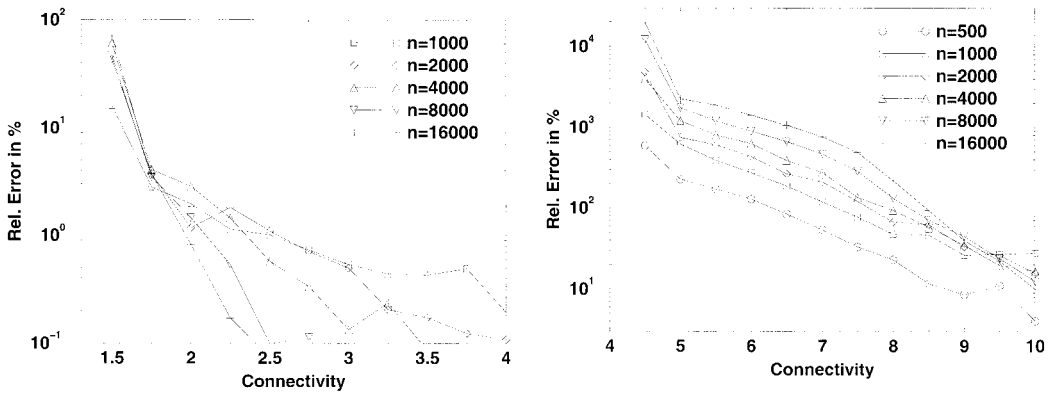
SPIN GLASSES

Of significant physical relevance are the low temperature properties of spin glasses [9], which we used to introduce EO above. With this implementation, we have studied the ground states of spin glasses in $d = 3$ and 4. Our data are discussed in detail in [12]. A fit of our data with $e_d(n) \sim e_d(\infty) + A/n$ for $n \to \infty$ predicts $e_3(\infty) = 1.7865(3)$ for $d = 3$ and $e_4(\infty) = 2.093(1)$ for $d = 4$. Both values are consistent with the findings of Refs. [22–24], providing independent confirmation of those results, with far less parameter tuning.

More recently [25] we have used EO to test theoretical predictions for the $T = 0$ properties of spin glasses on Bethe lattices [26]. Ref. [26] proposed a replica symmetry breaking (1RSB) solution for the ground state energy which significantly differs from replica symmetric (RS) results. As Figure 3 shows, the EO results are consistent with the 1RSB results and seem to rule out the RS solution.

Because EO never freezes into a local minimum, it is also well suited to enumerate the distribution of near-optimal states of a system, which is of great importance for the low-temperature dynamics in glassy materials [27,28]. Thus, we have also counted all lowest-energy states that EO could find for the Bethe lattice up to $n = 256$ (when memory became exhausted). The data extrapolate to a rather low

FIGURE 2

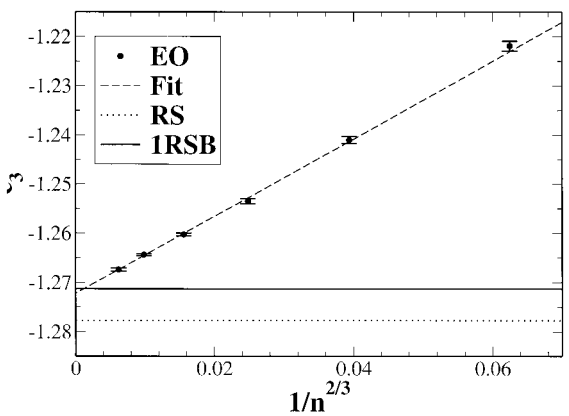


Plot of the error in the best result of SA relative to EO's on identical instances of random graphs (left) and geometric graphs (right) as function of the average connectivity c. The critical points for the GBP are at $c = 2 \ln 2 = 1.386$ for random and at $c \approx 4.5$ [11] for geometric graphs. SA's error near the critical point in both cases rises with n.

entropy per spin of $S/n = 0.009(2)$; there are no "free spins" [28] in an odd-connected lattice.

To gauge EO's performance for larger n, we have run our implementation also on two $3d$ lattice instances, *toruspm*3-8-50 and *toruspm*3-15-50, with $n = 8^3$ and $n = 15^3$, considered in the 7th DIMACS challenge for semi-definite problems [29]. Bounds [30] on the ground-state cost established for the larger instance are $C_{\text{lower}} = -6138.02$ (from semi-definite programming) and $C_{\text{upper}} = -5831$ (from branch-and-cut). EO found $C(S_{\text{best}}) = -6049$ (or $C/n = -1.7923$), a significant improvement on the upper bound an even lower than $e(\infty)$ from above. Furthermore, in that single run EO found 116 different states of that energy with Hamming-distances as far separated as 1500 mutually distinct spins! For the smaller instance the bounds given are -922 and -912, respectively, while EO finds -916 (or $C/n = -1.7891$). Here we found 10^5 such states before we terminated the run. Although this run (including sampling degenerate states!) took only a few minutes of CPU (at 800 MHz), the results for the larger instance require about 16 hours.

FIGURE 3



Plot of the extrapolation in n for the ground state energy e_3 on a 3-connected Bethe lattice. Assuming a fit with $e_3(n) \sim e_3(\infty) + a_1/n^{a_2}$ we obtain $e_3(\infty) = -1.2722(10)$, $a_1 = 0.80(5)$, and $a_2 = 0.67(1)$ with $\chi^2 = 3 \times 10^{-6}$, in good agreement with 1RSB predictions and well above the RS result.

ACKNOWLEDGMENTS

We thank the organizers of the workshop on "Complex Adaptive Systems" for their kind invitation, the HWK in Delmenhorst for its hospitality, and Emory University's URC and Los Alamos' LDRD program for financial support for this work.

REFERENCES

1. Bak, P. How Nature Works; Springer: New York, 1996.
2. Somfai, E.; Czirok, A.; Vicsek, T. Power-law distribution of landslides in an experiment on the erosion of a granular pile. J Phys A Math Gen 1994, 27, L757–L763.
3. Cieplak, M.; Giacometti, A.; Maritan, A.; Rinaldo, A.; Rodriguez-Iturbe, I.; Banavar, J.R. Models of fractal river basins. J Stat Phys 1998, 91, 1–15.
4. Paczuski, M.; Maslov, S.; Bak, P. Avalanche dynamics in evolution, growth, and depinning models. Phys Rev E 1996, 53, 414–443.
5. Bak, P.; Sneppen, K. Punctuated equilibrium and criticality in a simple model of evolution. Phys Rev Lett 1993, 71, 4083–4086.
6. Boettcher, S.; Paczuski, M. Exact results for spatiotemporal correlations in a self-organized critical model of punctuated equilibrium. Phys Rev Lett 1996, 76, 348.
7. Bak, P.; Tang, C.; Wiesenfeld, K. Self-Organized Criticality. Phys Rev Lett 1987, 59, 381.

8. Boettcher, S.; Percus, A.G. Nature's Way of Optimizing. Artif Intell 2000, 119, 275–286.
9. Mezard, M.; Parisi, G.; Virasoro, M.A. Spin Glass Theory and Beyond; World Scientific: Singapore, 1987.
10. Barahona, F. On the computational complexity of Ising spin glass models. J Phys A Math Gen 1982, 15, 3241–3253.
11. Boettcher, S. Extremal optimization and graph partitioning at the percolation threshold. J Math Phys A: Math Gen 1999, 32, 5201–5211.
12. Boettcher, S.; Percus, A.G. Optimization with extremal dynamics. Phys Rev Lett 2001, 86, 5211–5214.
13. Boettcher, S.; Percus, A.G. Extremal optimization for graph partitioning. Phys Rev E 2001, 64, 026114.
14. Boettcher, S.; Grigni, M. Jamming model for the extremal optimization heuristic. J Math Phys A Math Gen 2002, 35, 1109–1123.
15. Merz, P.; Freisleben, B. Memetic algorithms and the fitness landscape of the graph bipartitioning problem. Lect Notes Comput Sci 1998, 1498, 765–774.
16. Hendrickson, B.A.; Leland, R. A multilevel algorithm for partitioning graphs. In: Proceedings of Supercomputing '95; San Diego, CA, 1995.
17. Karypis, G.; Kumar, V. METIS, a Software Package for Partitioning Unstructured Graphs, see http://www-users.cs.umn.edu/karypis/metis/main.shtml (METIS is copyrighted by the Regents of the University of Minnesota).
18. Dall, J. Searching Complex State Spaces with Extremal Optimization and other Stochastic Techniques. Thesis, Fysisk Institut, Syddansk Universitet Odense, 2000 (in Danish).
19. Kirkpatrick, S.; Gelatt, C.D.; Vecchi, M.P. Optimization by simulated annealing. Science 1983, 220, 671–680.
20. Johnson, D.S.; Aragon, C.R.; McGeoch, L.A.; Schevon, C. Optimization by simulated annealing—an experimental evaluation. 1. Graph partitioning. Operations Res 1989, 37, 865–892.
21. Frontiers in problem solving: Phase transitions and complexity, Special issue of Artificial Intelligence 1996, 81, 1–2.
22. Pal, K.F. The ground state energy of the Edwards-Anderson Ising spin glass with a hybrid genetic algorithm. Physica A 1996, 223, 283–292.
23. Hartmann, A.K. Evidence for existence of many pure ground states in $3d \pm J$ spin glasses. Europhys Lett 1997, 40, 429.
24. Hartmann, A.K. Calculation of ground-state behavior of four-dimensional $\pm J$ Ising spin glasses. Phys Rev E 1999, 60, 5135–5138.
25. Boettcher, S. Numerical results for ground states of mean-field spin glasses at low connectivities. Phys Rev B, 2003, 67, 060403.
26. Mezard, M.; Parisi, G. The cavity method at zero temperature. http://xxx.lanl.gov/find/cond-mat/1/au:+Mezard_M/0/1/0/all/0/1; http://xxx.lanl.gov/find/cond-mat/1/au:+Parisi_G/0/1/0/all/0/1, J Stat Phys 2003, 111, 1.
27. Hartmann, A.K. Ground-state behavior of the three-dimensional $\pm J$ random-bond Ising model. Phys Rev B 1999, 59, 3617–3623.
28. Klotz, T.; Kobe, S. "Valley structures" in the phase space of a finite 3D Ising spin glass with $+$ or $-$ J interactions. J Phys A Math Gen 1994, 27, L95–L100.
29. 7th DIMACS Implementation Challenge on Semidefinite and related Optimization Problems; Johnson, D.S.; Pataki, G.; Alizadeh, F., Eds.; (to appear, see http://dimacs.rutgers.edu/Challenges/Seventh/#PC).
30. Jünger, M.; Liers, F., private communication.