Probabilistic Variational Formulation of Binary Programming

Arturo Berrones
Universidad Autónoma de Nuevo León, Facultad de Ingeniería Mecánica y Eléctrica, Posgrado en Ingeniería de Sistemas & Facultad de Ciencias Físico-Matemáticas, Posgrado en Ciencias con Orientación en Matemáticas, AP 126, Cd. Universitaria, San Nicolás de los Garza, NL 66450, México, arturo.berronessn@uanl.edu.mx

Jonás Velasco
CONACYT- Centro de Investigación en Matemáticas (CIMAT), A.C., Fray Bartolomé de las Casas 314, Barrio La Estación, CP 20259, Aguascalientes, Ags, México, jvelasco@cimat.mx

Juan Banda
Universidad Autónoma de Nuevo León, Facultad de Ingeniería Mecánica y Eléctrica, Posgrado en Ingeniería de Sistemas, AP 126, Cd. Universitaria, San Nicolás de los Garza, NL 66450, México, juan.bandamr@uanl.edu.mx

A probabilistic framework for large classes of binary integer programming problems is constructed. The approach is given by a mean field annealing scheme where the annealing phase is substituted by the solution of a dual problem that gives a lower (upper) bound for the original minimization (maximization) integer task. This bound has an information theoretic interpretation by which a principled feasible solution generator is constructed. The method is tested in linear and quadratic knapsack problems for which is capable to find high quality solutions in running times that are orders of magnitude shorter than state of the art algorithms. Experimental evidence indicates that for the quadratic case, the mean field approximation improves with problem size for unstructured instances. This is reminiscent of the exact mean field limit found in several spin glass models.

Key words: heuristics; stochastic search; binary optimization

1. Introduction

Heuristics can be interpreted as solution generators that involve diversification and intensification mechanisms. Diversification usually introduces randomness to the search of solutions and intensification gives strategies to adapt the solutions to the optimization task at hand. Is proved in this contribution that for the class of binary constrained

programs, the general scheme of heuristics can be formally expressed by the variational construction of a probabilistic model. Adaptation arises from averaging the binary decision variables, an operation that translates the original task into a continuous variational problem. Diversification do not involves explicit random processes but the optimization of a suitable information measure given in terms of the dual variables associated to the original constraints. Meaningful bounds for the optimal value of the binary task follow. In the case of linear binary programs the variational problem can be analytically tackled leading to an overall polynomial complexity. Is numerically shown that the probabilistic framework gives high quality solutions in short computation times even in the context of binary problems that are NP-hard in the strong sense. The presented ideas are general enough to be extended to integer optimization in bounded domains.

2. Dual mean field annealing

Our framework is based on mean field annealing, a version of simulated annealing originally proposed to deal with combinatorial optimization problems in which the sampling stage is replaced by a step in which a set of coupled self-consistent field equations must be solved at each temperature (randomness parameter) T, which reduces much of the computational burden of simulated annealing, see Ohlsson and Pi (1995), Karimi and AziziVerki (2012) and Ohlsson et al. (2001). We demonstrate that under the basis of mean field annealing, the principles of relaxation and duality can be invoked to substitute the annealing phase by the solution of a dual problem which gives a lower (upper) bound for the original minimization (maximization) integer task. To proceed, consider the following class of optimization problems,

$$\begin{aligned} &\min f(\vec{x}) \quad s.t. \\ &g_k(\vec{x}) \leq 0, h_l(\vec{x}) = 0, \end{aligned} \tag{1}$$

where $\vec{x}$ is a vector of binary decision variables, $g_{k(k=1,\ldots,K)}$ are inequality constraint functions and $h_{l(l=1,\ldots,L)}$ are equality constraints. The optimization task (1) can in principle be represented by a potential function $V(\vec{x})$ which includes the objective and the

constraints. A probability distribution can be associated to such a potential by the transformation, see Kirkpatrick et al. (1983) and Metropolis et al. (1953),

$$P(\vec{x}) = \frac{1}{Z} \exp\left(-V/T\right), \tag{2}$$

where Z is a normalization factor (or partition function). The Eq. (2) gives the maximum entropy distribution which is consistent with the condition $\langle V \rangle_P = \int V P d\vec{x}$, see Jaynes (1957). However, P is in general intractable. Moreover, in our setup is not even known, because the explicit definition of V would require the knowledge of suitable "barrier" terms that exactly represent the constraints. Is therefore proposed the following mean field probabilistic model for the decision variables,

$$Q(\vec{x}) = \prod_{i=1}^{N} p(x_i), \tag{3}$$

Mean field techniques, together with other methods which have first emerged in statistical mechanics, see Parisi (1988), have been already successfully applied to discover fundamental features of combinatorial problems and valuable solution strategies, see Martin et al. (2001), Hogg (1996), Ohlsson and Pi (1995), Karimi and AziziVerki (2012) and Ohlsson et al. (2001). Our purpose in this contribution is to develop a mean field framework to find good candidate solutions to linear and nonlinear binary problems in the constrained situation (1) by the use of the principles of relaxation and duality. The most general form for the independent marginals is,

$$p(x_i) = 1 + (2m_i - 1)x_i - m_i. \tag{4}$$

The m's are continuous mean field parameters, $m \in [0, 1]$. These parameters can be selected by the minimization of the Kullback-Leibler divergence between distributions Q and P, see Opper and Saad (2001),

$$D_{KL}(Q||P) = \langle \ln Q \rangle - \langle \ln P \rangle, \tag{5}$$

where the brackets represent averages with respect to the tractable distribution Q. Introducing the entropy $S_Q = -T\langle \ln Q \rangle$, the variational problem $\min F_Q$ is obtained, where

$$F_Q = \left[\frac{1}{T} \langle V \rangle - S_Q \right] \tag{6}$$

is the variational "free energy" of the distribution Q, see Opper and Saad (2001). In first instance we consider the class of combinatorial optimization problems in which all the involved functions (objective and constraints) are polynomial, e. g. $f(\vec{x}) = a_o + \sum_i b_i x_i + \sum_i \sum_j c_{i,j} x_i x_j + \sum_i \sum_j \sum_r q_{i,j,r} x_i x_j x_r + ...$ In such case $\langle f(\vec{x}) \rangle = f(\langle \vec{x} \rangle)$, $\langle g(\vec{x}) \rangle = g(\langle \vec{x} \rangle)$ and $\langle h(\vec{x}) \rangle = h(\langle \vec{x} \rangle)$. Therefore, the continuous relaxation of the problem (1) is equivalent to its average under the mean field distribution,

$$\begin{aligned} \min f(\vec{m}) \quad s.t. \\ g_k(\vec{m}) \leq 0, h_l(\vec{m}) = 0. \end{aligned} \tag{7}$$

An expression for $\langle V \rangle$ can be constructed in terms of the Lagrangian,

$$\mathcal{L} = f(\vec{m}) + \sum_l \lambda_l h_l(\vec{m}) + \sum_k \mu_k g_k(\vec{m}), \tag{8}$$

where the parameters λ_l and $\mu_k \geq 0$ are the Karush-Kuhn-Tucker (KKT) generalization of the Lagrange multipliers, see Chong and Zak (2013). The entropy of Q, on the other hand, is given by,

$$S_Q = -T \sum_i \left[(1 - m_i) \ln(1 - m_i) + m_i \ln m_i \right], \tag{9}$$

so the variational problem for the m's is written like,

$$\begin{aligned} \min F_Q(\vec{m}) = \min \quad & \frac{1}{T} \left[f(\vec{m}) + \sum_l \lambda_l h_l(\vec{m}) + \sum_k \mu_k g_k(\vec{m}) \right] \\ & + T \sum_i \left[(1 - m_i) \ln(1 - m_i) + m_i \ln m_i \right]. \end{aligned} \tag{10}$$

Equations (3), (4) and (10) give a general probabilistic model for combinatorial optimization problems with binary decision variables. Any continuous and differentiable nonlinearities in the objective or the constraints can be expanded in a Taylor series under to the condition $m_i < 1 \ \forall \ i$. Due to independence under the mean field, $\langle V \rangle$ is therefore given by Eq. (8) for any problem (1), provided that the stated conditions are met. Stationarity applied to F_Q with respect to the mean field parameters reduce the variational problem to a set of self-consistency equations for $\vec{m}$ at fixed values of the multipliers,

$$m_i = \frac{1}{1 + \exp\left[\frac{\partial_i \mathcal{L}_{\vec{\lambda},\vec{\mu}}(\vec{m})}{T^2}\right]}. \tag{11}$$

If on the other hand the m's are fixed, a dual problem can be defined in terms of the λ's and μ's which in this case are interpreted as a set of dual variables. The analysis of this dual problem leads to meaningful bounds for the optimum of the original constrained task. Consider the operator $round(\vec{m})$ which associates an integer to each component of $\vec{m}$. If x_o is the minimum of (1) and $\vec{x} = round(\vec{m})$ is a vector that satisfies the constraints, at $T = 1$ follows that

$$\begin{aligned} f(\vec{x}_o) &\leq f(\vec{x}), \\ F_Q(\vec{m}) &\leq f(\vec{m}) \end{aligned} \tag{12}$$

and then

$$F_Q(\vec{m}) \leq f(\vec{x}_o) \leq f(\vec{x}). \tag{13}$$

Therefore, the solution of

$$\max_{\{\vec{\lambda},\vec{\mu}\}} \quad \left[\min_{\{\vec{m}\}} \quad F_Q(\vec{m})\right] \tag{14}$$

maximizes a lower bound on $f(\vec{x}_o)$ which minimizes the Kullback-Leibler divergence between distributions Q and P. By inserting the optimal $\vec{\lambda}$ into Eq. (11), the distribution (4) is the best possible tractable, i. e. that can be efficiently sampled, probabilistic model from which vectors close to the solution of the actual binary task can be drawn.

3. Linear binary optimization

For linear objective and linear inequality constraints, dual mean field annealing reduces to an analytically amenable variational problem for the m's and a linear optimization task for the dual variables. Consider the binary optimization problem,

$$\begin{aligned} \min \quad & -\vec{q}\cdot\vec{x} \quad s.t. \\ & \vec{w}_1\cdot\vec{x}-d_1\leq 0, \\ & \vec{w}_2\cdot\vec{x}-d_2\leq 0, \\ & \ldots \\ & \vec{w}_k\cdot\vec{x}-d_k\leq 0, \\ & \vec{w}_K\cdot\vec{x}-d_K\leq 0. \end{aligned} \tag{15}$$

The solution of the variational problem in this case reads,

$$m_i=\frac{1}{1+\exp(-q_i+\sum_{k=1}^{K}\mu_k w_{k,i})}. \tag{16}$$

At any fixed value $\vec{m}=\vec{m}_o$ the free energy function is clearly linear in $\vec{\mu}$ and the maximum value of the free energy in the dual space occurs in one of the vertices of a hypercube defined by $\vec{\mu}_{start}=\vec{0}$ and $\vec{\mu}_{end}$ values of the multipliers. It's always possible to find large enough values for $\vec{\mu}_{end}$ such that $\vec{x}=round(\vec{m})$ satisfies the constraints. The conclusion is that the exhaustive exploration of the dual space to maximize F_Q can be done by performing 2^{K-1} line searches, each one corresponding to one of the diagonals of the hypercube.

4. Linear or nonlinear objective function and a single linear constraint

We have implemented the procedure given in Algorithm 1 for problems with a single linear constraint. The method essentially consists on a bisection search on the dual space that maximizes the free energy. In the pseudocode, μ_s and μ_f define an interval for the

Algorithm 1 (Pseudo-code of the Dual Mean Field –DMF– algorithm)

Initialize: $\mu_s \leftarrow 0$, $\mu_f \leftarrow \mu_o$
Set $\epsilon \leftarrow \infty$, $TOL \leftarrow 0.01$, $k \leftarrow 1$
Define $m_i^* \leftarrow 1$ and $x_i^{\text{best}} \leftarrow 0$, for all $i = 1, ..., N$
Compute $\vec{m}$ in Eq. (11) with $\mu \leftarrow \mu_f$ and $\vec{m}^*$
Define $\vec{x} \leftarrow round(\vec{m})$
Compute feasibility of $\vec{x}$
while $\vec{x}$ is infeasible **do**
    $\mu_s \leftarrow \mu_f$
    $\mu_f \leftarrow 3 * \mu_f$
    Compute $\vec{m}$ in Eq. (11) with $\mu \leftarrow \mu_f$ and $\vec{m}^*$
    Define $\vec{x} \leftarrow round(\vec{m})$
    Compute feasibility of $\vec{x}$
end while
while $(\epsilon > TOL)$ **do**
    $\mu_k \leftarrow (\mu_s + \mu_f)/2$
    Compute $\vec{m}_k$ in Eq. (11) with $\mu \leftarrow \mu_k$ and $\vec{m}^*$
    Define $\vec{x}_k \leftarrow round(\vec{m}_k)$
    Calculate feasibility of $\vec{x}_k$
    if $\vec{x}_k$ is infeasible **then**
        $\mu_s \leftarrow \mu_k$
    else
        $\mu_f \leftarrow \mu_k$
        $\vec{x}_{\text{best}} \leftarrow \vec{x}_k$
    end if
    $k \leftarrow k + 1$
    $\epsilon \leftarrow \mu_f - \mu_s$
end while
Output best solution $\vec{x}_{\text{best}}$ found

multiplier value. Initally the lower limit is taken to zero and the upper limit is given by a reasonable guess for the relative importance of the constraint, such that the distribution (4) actually represents the "competition" between the objective and the constraints under small changes in the solution vector. It should be however noted that is always possible to give a sufficiently high initial value such that the constraint is satisfied. Due to the linearity of the constraint, the bisection method converges to the minimum value of the multiplier that gives a feasible solution after rounding.

5. Numerical experiments

5.0.1. Classical knapsack problem The probabilistic setup (3), (4), (10) and (11) is now tested on specific examples. At first instance the classical Knapsack Problem (KP) is considered,

$$\begin{aligned} \min \quad & -\vec{q}\cdot\vec{x} \quad s.t. \\ & \vec{w}\cdot\vec{x} - d \leq 0, \end{aligned} \tag{17}$$

where d is the capacity of the knapsack, $\vec{q}$ are the gains and $\vec{w}$ the weights of a collection of $i = 1, ..., N$ objects. The classical Knapsack Problem is important from both theoretical and practical standpoints. In its decision version, is an easy to formulate NP problem and therefore offers an ideal playground for the study of computational complexity issues, see Kellerer et al. (2004). It's also the basis for models that arise in applications such like resource allocation, see Luss (2012), optimal portfolios, see Roland et al. (2016) and planning, see Dolgui et al. (2015), among many others, see Chhajed (2008).

To test our dual mean field method in the classical KP, Algorithm 1 is initialized with $\mu_f = \frac{\overline{q}}{\overline{w}}$, where the overlines represent the mean values of the instance parameters. This initial guess for μ_f can be interpreted like the dual value at which the constraint and the objective remain comparable under small changes in the solution vector. We have generated instances with the Pisinger generator [1], see Pisinger (2005), under the

[1] `http://www.diku.dk/~pisinger/generator.c`

conditions of strong linear correlations between q_i and w_i, being w_i randomly distributed in the interval $[1, R]$ and $q_i = w_i + R/10$. Due to its simple structure, large instances of KP can be solved exactly and quite efficiently even for large problem sizes at moderate R values, but at large values of R most algorithms encounter difficulties with large instances, see Pisinger (2005). To illustrate this, we consider here ten instances of each of the sizes $10^2, 5\times10^2, 10^3, 5\times10^3, 10^4, 5\times10^4, 10^5, 5\times10^5, 10^6, 5\times10^6, 10^7$ and 5×10^7 with R values in the range $[10^2, 10^7]$, for which we have applied a state of the art implementation of branch and bound provided by the advanced commercial solver Cplex [2], version 12.6.2. Because for problem sizes above 5000 with $R \geq 10^3$ the computing time to find the optimal solution becomes prohibitively large in our considered equipment (workstation with an Intel® Xeon® 3.40GHz $\times$ 8 processor, 16 GB RAM and Linux Ubuntu 14.04 LTS operating system), branch and bound is treated like a heuristic: the procedure stops when the first feasible solution is found. The average times obtained by Cplex branch and bound are shown in Figure 1. In Figure 2 the same is reported for our dual mean field annealing. While the exponential time explosion is evident for branch and bound, which is unable to find any feasible solution in problems with size, $N > 10^6$ with $R = 10^3$, $N > 5\times10^5$ with $R = 10^4$ or 10^5, and $N > 5\times10^4$ with $R = 10^6$ or 10^7, dual mean field annealing finds feasible solutions in all the considered instances at very low computation times. Moreover, these are high quality solutions, which is illustrated in Figure 3. State of the art heuristics that deal with large scale linear KP report solutions of problems up to 10^5 and $R \leq 2000$, see Banitalebi et al. (2016), Boyer (2012) and Kong et al. (2015). Our interest with KP is to show that dual mean field annealing behaves in a robust way with respect to R, giving valuable solutions for ranges well beyond the scope of state of the art exact or heuristic methods.

[2] `http://www-01.ibm.com/software/commerce/optimization/cplex-optimizer/index.html`

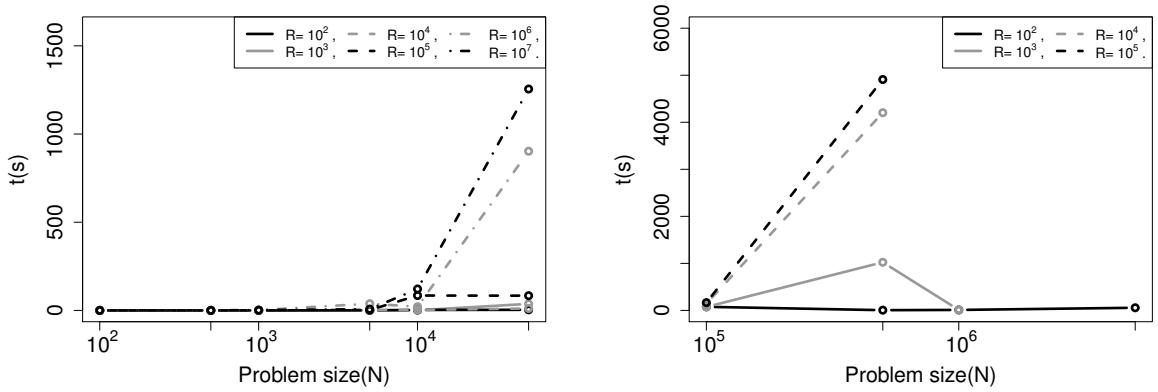

Figure 1 **The average times obtained by Cplex branch and bound for the classical Knapsack Problem.**

5.0.2. Quadratic knapsack problem The effectivity of the formulation in nonlinear cases has been tested on the Quadratic Knapsack Problem (QKP), which is stated as follows, see Pisinger (2007),

$$\begin{aligned} \min \quad & -\vec{x}^t \mathcal{Q} \vec{x} \quad s.t. \\ & \vec{w} \cdot \vec{x} - d \leq 0, \end{aligned} \tag{18}$$

where $\mathcal{Q}$ is a symmetric matrix with coefficients $q_{i,j} \geq 0 \; \forall \; i, j$. QKP has a graph-theoretic interpretation in terms of the *Clique* problem, see Pisinger (2007). Moreover, QKP is NP-hard in the strong sense, see Pisinger (2007). To our knowledge, QKP has not been previously studied by mean field annealing techniques. The mean field Lagrangian reads,

$$\mathcal{L} = -\vec{m}^t \mathcal{Q} \vec{m} + \mu(\vec{w} \cdot \vec{m} - d), \tag{19}$$

from which

$$\partial_i \mathcal{L}_\mu = -2q_{i,i}m_i - \sum_{j \neq i} q_{i,j} m_j + \mu w_i. \tag{20}$$

The initial value of the multiplier is taken like,

$$\mu = \frac{1}{N\bar{w}} \left[2 \sum_i q_{i,i} + \sum_i \sum_{j \neq i} q_{i,j} \right]. \tag{21}$$

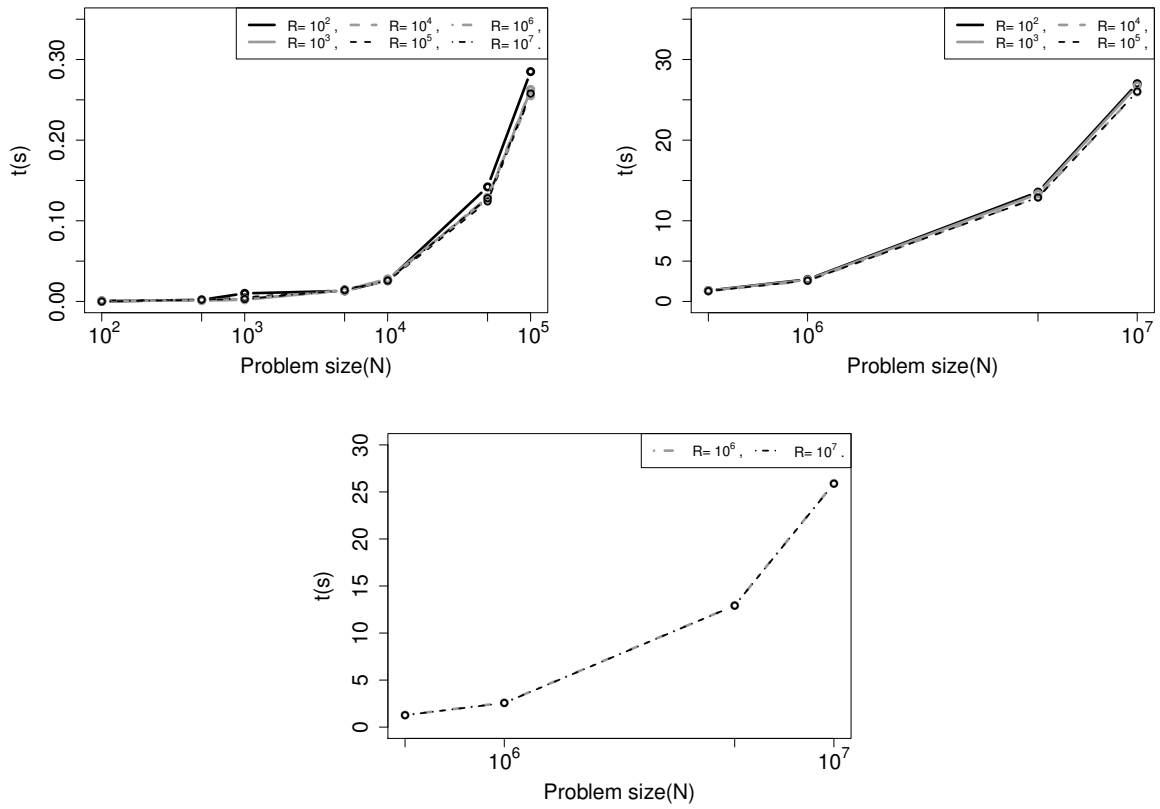

Figure 2 The average times obtained by dual mean field annealing for the classical Knapsack Problem.

Obtaining exact solutions is extremely difficult for the QKP, even for some hundreds of variables, see Billionnet and Soutif (2004) and Fomeni and Letchford (2013). One of the few existing benchmarks with known exact optima is provided by Billionnet and Soutif (2004) [3] [4] , which considers system sizes of 100, 200 and 300 with varying $\mathcal{Q}$ matrix densities. The behavior of dual mean field annealing in these relatively small instances is shown in Table 1 and summarized in Figure 4. The average gap is below 3.5% for all system sizes. Interestingly, the gaps are smaller for the larger sizes. To assess the performance of dual mean field annealing for increasing system size in comparison

[3] `http://cedric.cnam.fr/~soutif/QKP/QKP.html`

[4] Instances 100_100_4, 200_25_3 and 300_25_3 are not available for QKP, they are not considered.

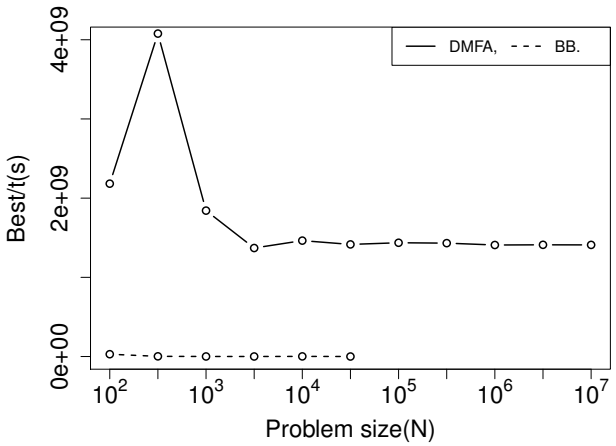

Figure 3 **Comparison at $R = 10^7$ of the ratios between the best solution values and the computing times between Cplex branch and bound and dual mean field annealing.**

with state of the art heuristics, we have considered the benchmark by Yang et al. (2013), which consists of problems with 1000 and 2000 variables with 40 instances for each size. No exact optima is known for these instances. The benchmark has been studied by a dynamic programming heuristic, see Fomeni and Letchford (2013), a GRASP with Tabu Search, see Yang et al. (2013) and an iterated hyperplane exploration, see Chen and Hao (2017). The complete comparisons against dual mean field are presented in Tables 2 and 3 and summarized in Figures 5 and 6. The best objective values and associated computing times for the methods other than our mean field framework are taken from Chen and Hao (2017). The computing times across different equipments is normalized according to the formula, see Yuan et al. (2009):

$$\text{scaled CPU time(s)} = \frac{\text{given CPU speed(GHz)}}{2.8\ \text{GHz}} \times \text{given CPU time(s)} \tag{22}$$

Dual mean field annealing obtains competitive solutions in computing times that are orders of magnitude shorter than the most advanced heuristics for this benchmark. We have considered an additional set of very large instances (sizes 5000 and 6000) due to Chen and Hao (2017). To our knowledge, the only previous heuristic capable to deal with

this kind of QKP sizes is the iterated hyperplane exploration, against which dual mean field is compared in Table 4 and Figure 7 and 8. Again, dual mean field gives high quality solutions in computing times that are orders of magnitude shorter. The gaps of the dual mean field solutions with respect to the best solutions known for all of the considered QKP instances is reported in Figure 9. Interestingly, the gaps diminish with problem size. This behavior is consistent with very well known results of mean field theories in physics, where the mean field approximations increase in accuraccy as the dimension of nonlinear interacting systems grows, see Sherrington et al. (1975) and Thouless et al. (1977).

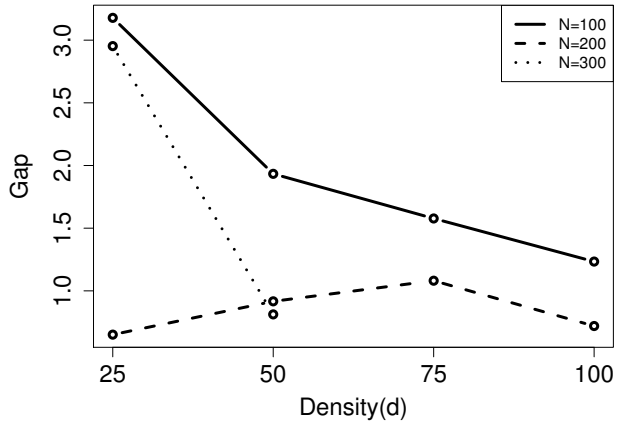

Figure 4 **Gaps obtainded by DMFA on the 100 small and medium sized benchmark instances generated by Billionnet and Soutif (2004). The gaps are smaller for the larger sizes.**

6. Discussion: exact mean field limit in spin glasses and its implications for the quadratic knapsack problem

The experiments presented in the last section point to an interesting analogy between the behavior of spin glasses and QKP under the mean field approximation. A spin glass is a model from statistical physics that describes general properties of amorphous materials,

Table 1 Results of DMFA on the 100 small and medium sized benchmark instances generated by Billionnet and Soutif (2004).

d_l	Opt	N=100			*d_l*	Opt	N=200			*d_l*	Opt	N=300		
		Best	Gap	t(s)			Best	Gap	t(s)			Best	Gap	t(s)
25_1	18,558	17,341	6.56	0.0003	25_1	204,441	202,122	1.13	0.0010	25_1	29,140	26,835	7.91	0.0033
25_2	56,525	56,339	0.33	0.0004	25_2	239,573	239,287	0.12	0.0009	25_2	281,990	278,475	1.25	0.0021
25_3	3,752	3,596	4.16	0.0005	25_3	245,463	–	–	–	25_3	231,075	–	–	–
25_4	50,382	48,785	3.17	0.0005	25_4	222,361	220,296	0.93	0.0010	25_4	444,759	442,569	0.49	0.0021
25_5	61,494	60,983	0.83	0.0004	25_5	187,324	185,000	1.24	0.0010	25_5	14,988	14,401	3.92	0.0021
25_6	36,360	35,162	3.29	0.0006	25_6	80,351	77,406	3.67	0.0010	25_6	269,782	268,686	0.41	0.0016
25_7	14,657	14,221	2.97	0.0004	25_7	59,036	57,884	1.95	0.0011	25_7	485,263	482,847	0.50	0.0018
25_8	20,452	19,317	5.55	0.0003	25_8	149,433	147,990	0.97	0.0010	25_8	9,343	8,440	9.66	0.0017
25_9	35,438	34,535	2.55	0.0005	25_9	49,366	47,693	3.39	0.0010	25_9	250,761	246,879	1.55	0.0017
25_10	24,930	24,340	2.37	0.0002	25_10	48,459	46,109	4.85	0.0010	25_10	383,377	379,966	0.89	0.0017
50_1	83,742	82,959	0.94	0.0003	50_1	372,097	371,231	0.23	0.0009	50_1	513,379	509,643	0.73	0.0017
50_2	104,856	104,389	0.45	0.0007	50_2	211,130	209,332	0.85	0.0010	50_2	105,543	104,980	0.53	0.0019
50_3	34,006	32,220	5.25	0.0004	50_3	227,185	223,330	1.70	0.0010	50_3	875,788	874,538	0.14	0.0019
50_4	105,996	105,640	0.34	0.0004	50_4	228,572	224,544	1.76	0.0010	50_4	307,124	304,354	0.90	0.0020
50_5	56,464	56,010	0.80	0.0008	50_5	479,651	477,627	0.42	0.0010	50_5	727,820	726,850	0.13	0.0019
50_6	16,083	15,038	6.50	0.0005	50_6	426,777	424,724	0.48	0.0010	50_6	734,053	732,855	0.16	0.0018
50_7	52,819	51,979	1.59	0.0008	50_7	220,890	218,224	1.21	0.0010	50_7	43,595	41,854	3.99	0.0017
50_8	54,246	53,850	0.73	0.0005	50_8	317,952	316,732	0.38	0.0010	50_8	767,977	764,955	0.39	0.0017
50_9	68,974	68,082	1.29	0.0003	50_9	104,936	103,494	1.37	0.0010	50_9	761,351	758,314	0.40	0.0018
50_10	88,634	87,347	1.45	0.0003	50_10	284,751	282,609	0.75	0.0009	50_10	996,070	988,765	0.73	0.0017
75_1	189,137	189,137	0.00	0.0004	75_1	442,894	437,929	1.12	0.0010					
75_2	95,074	93,532	1.62	0.0006	75_2	286,643	283,839	0.98	0.0011					
75_3	62,098	61,806	0.47	0.0007	75_3	61,924	60,698	1.98	0.0011					
75_4	72,245	70,576	2.31	0.0007	75_4	128,351	127,050	1.01	0.0011					
75_5	27,616	26,950	2.41	0.0005	75_5	137,885	134,814	2.23	0.0009					
75_6	145,273	144,731	0.37	0.0003	75_6	229,631	228,765	0.38	0.0010					
75_7	110,979	109,214	1.59	0.0003	75_7	269,887	268,312	0.58	0.0010					
75_8	19,570	18,726	4.31	0.0004	75_8	600,858	600,631	0.04	0.0010					
75_9	104,341	102,313	1.94	0.0002	75_9	516,771	514,428	0.45	0.0010					
75_10	143,740	142,676	0.74	0.0002	75_10	142,694	139,787	2.04	0.0011					
100_1	81,978	81,760	0.27	0.0003	100_1	937,149	935,422	0.18	0.0009					
100_2	190,424	188,993	0.75	0.0007	100_2	303,058	300,427	0.87	0.0010					
100_3	225,434	225,124	0.14	0.0004	100_3	29,367	28,455	3.11	0.0011					
100_4	63,028	–	–	–	100_4	100,838	99,610	1.22	0.0010					
100_5	230,076	224,885	2.26	0.0006	100_5	786,635	785,980	0.08	0.0012					
100_6	74,358	73,855	0.68	0.0005	100_6	41,171	41,171	0.00	0.0011					
100_7	10,330	9,859	4.56	0.0007	100_7	701,094	698,219	0.41	0.0009					
100_8	62,582	61,858	1.16	0.0004	100_8	782,443	779,304	0.40	0.0010					
100_9	232,754	231,760	0.43	0.0002	100_9	628,992	626,879	0.34	0.0011					
100_10	193,262	191,578	0.87	0.0004	100_10	378,442	376,219	0.59	0.0009					

see Young (1997). In its simplest form, a collection of *spins* (elementary particles with intrinsic magnetic moment) interact pairwise to give a total magnetic energy,

$$E = -\sum_{(i,j)} J_{i,j} x_j x_i, \tag{23}$$

Table 2 Comparative results of DMFA with 3 state-of-the-art algorithms on the 40 large-sized (N=1000) benchmark instances generated by Yang et al. (2013).

N_d_l	DMFA		DP+FE (Fomeni and Letchford 2013)		GRASP+Tabu (Yang et al. 2013)		IHEA (Chen and Hao 2017)	
	Best	t(s)	Best	t(s)	Best	t(s)	Best	t(s)
1000_25_1	6,150,753	0.017	6,172,407	1,682.280	6,172,407	18.234	6,172,407	2.765
1000_25_2	224,731	0.016	229,833	2,103.290	229,941	20.448	229,941	5.390
1000_25_3	167,473	0.017	172,418	1,919.350	172,418	13.429	172,418	5.892
1000_25_4	359,192	0.016	367,365	2,537.720	367,426	16.188	367,426	7.293
1000_25_5	4,872,931	0.016	4,885,569	2,626.970	4,885,611	23.368	4,885,611	5.543
1000_25_6	13,361	0.017	15,528	608.550	15,689	5.072	15,689	1.635
1000_25_7	4,940,406	0.016	4,945,741	2,725.220	4,945,810	22.636	4,945,810	4.804
1000_25_8	1,699,843	0.016	1,709,954	3,762.890	1,710,198	44.150	1,710,198	7.104
1000_25_9	489,517	0.016	496,315	2,839.990	496,315	18.619	496,315	6.891
1000_25_10	1,164,491	0.016	1,173,686	3,607.270	1,173,792	36.537	1,173,792	7.573
1000_50_1	5,652,232	0.016	5,663,517	3,722.470	5,663,590	31.459	5,663,590	6.870
1000_50_2	178,332	0.016	180,831	1,450.870	180,831	0.893	180,831	3.692
1000_50_3	11,363,521	0.017	11,384,139	2,071.250	11,384,283	19.753	11,384,283	3.338
1000_50_4	317,240	0.017	322,184	1,868.860	322,226	13.677	322,226	5.433
1000_50_5	9,967,800	0.016	9,983,477	2,570.760	9,984,247	25.315	9,984,247	3.662
1000_50_6	4,092,241	0.018	4,106,186	3,801.720	4,106,261	36.010	4,106,261	7.691
1000_50_7	10,493,537	0.016	10,498,135	2,322.160	10,498,370	20.727	10,498,370	3.584
1000_50_8	4,975,984	0.018	4,981,017	3,826.980	4,981,146	72.100	4,981,146	9.155
1000_50_9	1,720,453	0.016	1,727,727	3,382.020	1,727,861	32.717	1,727,861	9.381
1000_50_10	2,329,531	0.017	2,340,590	3,605.070	2,340,724	59.074	2,340,724	7.416
1000_75_1	11,554,653	0.017	11,569,498	3,334.210	11,570,056	39.680	11,570,056	4.892
1000_75_2	1,894,859	0.017	1,901,119	3,094.560	1,901,389	20.131	1,901,389	6.492
1000_75_3	2,092,412	0.017	2,096,415	3,208.980	2,096,485	24.713	2,096,485	8.742
1000_75_4	7,293,839	0.017	7,305,195	3,821.020	7,305,321	34.156	7,305,321	6.846
1000_75_5	13,950,705	0.017	13,969,705	2,887.190	13,970,240	23.182	13,970,842	6.022
1000_75_6	12,274,739	0.017	12,288,299	3,178.950	12,288,738	20.733	12,288,738	4.463
1000_75_7	1,092,797	0.018	1,095,837	2,580.270	1,095,837	14.359	1,095,837	7.119
1000_75_8	5,564,147	0.016	5,575,592	3,804.420	5,575,813	42.451	5,575,813	7.833
1000_75_9	687,813	0.017	695,595	2,171.330	695,774	14.062	695,774	4.624
1000_75_10	2,501,816	0.017	2,507,627	3,349.440	2,507,677	29.338	2,507,677	6.863
1000_100_1	6,231,812	0.016	6,243,330	3,849.500	6,243,494	44.646	6,243,494	7.018
1000_100_2	4,837,713	0.017	4,853,927	3,627.050	4,854,086	52.601	4,854,086	7.092
1000_100_3	3,157,656	0.017	3,171,955	3,320.520	3,172,022	29.177	3,172,022	6.391
1000_100_4	749,978	0.017	754,542	1,990.800	754,727	14.651	754,727	5.207
1000_100_5	18,626,930	0.017	18,646,607	2,829.350	18,646,620	24.273	18,646,620	4.070
1000_100_6	16,004,630	0.017	16,019,697	3,247.810	16,018,298	25.780	16,020,232	5.204
1000_100_7	12,933,578	0.017	12,936,205	3,587.160	12,936,205	27.590	12,936,205	5.533
1000_100_8	6,887,903	0.017	6,927,342	3,850.890	6,927,738	59.551	6,927,738	7.298
1000_100_9	3,870,342	0.017	3,874,959	3,463.920	3,874,959	32.414	3,874,959	7.085
1000_100_10	1,328,223	0.016	1,334,389	2,474.890	1,334,494	14.651	1,334,494	6.270

where the x's represent the spins and the $J_{i,j}$'s are the exchange interactions. Equation (23) is essentially the quadratic knapsack objective function. In physics it's of interest to find the minumum energy configurations of the glass at zero temperature (a problem equivalent to an unconstrained QKP) or to describe the structural phases of the system at different temperatures, see Opper and Saad (2001) and Young (1997). A mean field description of (23) becomes exact if the spin values don't fluctuate: if we write $x_i = m_i + \delta_i$ $\forall\ i$, the average energy is given by $\langle E \rangle = -\sum_{(i,j)} J_{i,j}\left[m_j m_i - \langle \delta_i \delta_j \rangle\right]$. If the fluctuations

Table 3 Comparative results of DMFA with 3 state-of-the-art algorithms on the 40 large-sized (N=2000) benchmark instances generated by Yang et al. (2013).

N_d_l	DMFA		DP+FE (Fomeni and Letchford 2013)		GRASP+Tabu (Yang et al. 2013)		IHEA (Chen and Hao 2017)	
	Best	t(s)	Best	t(s)	Best	t(s)	Best	t(s)
2000_25_1	5,251,589	0.071	5,268,004	57,726.920	5,268,188	320.273	5,268,188	22.264
2000_25_2	13,270,768	0.070	13,293,940	51,050.130	13,294,030	205.053	13,294,030	24.917
2000_25_3	5,477,886	0.071	5,500,323	57,419.270	5,500,433	496.081	5,500,433	28.933
2000_25_4	14,606,794	0.070	14,624,769	46,620.160	14,625,118	215.072	14,625,118	17.050
2000_25_5	5,957,249	0.070	5,975,645	57,416.960	5,975,751	457.765	5,975,751	28.102
2000_25_6	4,474,141	0.071	4,491,533	56,155.800	4,491,691	294.252	4,491,691	23.442
2000_25_7	6,372,821	0.070	6,388,475	57,116.940	6,388,756	346.090	6,388,756	25.178
2000_25_8	11,746,351	0.069	11,769,395	52,832.060	11,769,873	277.109	11,769,873	22.584
2000_25_9	10,941,184	0.070	10,959,388	54,258.650	10,960,328	278.882	10,960,328	22.420
2000_25_10	134,917	0.070	139,233	14,686.960	139,236	68.070	139,236	7.551
2000_50_1	7,053,129	0.069	7,070,736	52,860.690	7,070,736	294.078	7,070,736	28.016
2000_50_2	12,562,364	0.070	12,586,693	57,518.440	12,587,545	331.619	12,587,545	23.943
2000_50_3	27,246,961	0.070	27,266,846	48,397.300	27,268,336	191.506	27,268,336	22.691
2000_50_4	17,720,044	0.070	17,754,391	57,376.090	17,754,434	485.249	17,754,434	24.506
2000_50_5	16,773,712	0.070	16,804,699	57,563.580	16,805,490	923.936	16,806,059	32.057
2000_50_6	23,047,769	0.070	23,075,693	52,613.210	23,076,155	285.256	23,076,155	21.579
2000_50_7	28,753,633	0.069	28,757,657	46,437.960	28,759,759	442.792	28,759,759	25.365
2000_50_8	1,567,412	0.070	1,580,242	32,416.870	1,580,242	102.412	1,580,242	13.937
2000_50_9	26,492,874	0.070	26,523,637	48,529.930	26,523,791	212.114	26,523,791	19.695
2000_50_10	24,727,531	0.069	24,746,249	50,565.420	24,747,047	253.202	24,747,047	20.613
2000_75_1	25,098,111	0.070	25,121,327	57,579.990	25,121,998	500.371	25,121,998	22.721
2000_75_2	12,635,538	0.070	12,663,927	54,629.120	12,664,670	316.231	12,664,670	21.584
2000_75_3	43,928,427	0.070	43,943,294	45,151.420	43,943,994	171.362	43,943,994	18.723
2000_75_4	37,476,687	0.070	37,496,414	50,255.520	37,496,613	219.561	37,496,613	19.901
2000_75_5	24,793,604	0.070	24,835,254	56,840.030	24,834,948	424.285	24,835,349	27.439
2000_75_6	45,127,942	0.069	45,137,702	44,437.730	45,137,758	190.011	45,137,758	20.862
2000_75_7	25,486,552	0.068	25,502,503	57,480.680	25,502,608	303.887	25,502,608	21.848
2000_75_8	10,035,317	0.070	10,067,752	52,566.820	10,067,892	213.795	10,067,892	21.560
2000_75_9	14,173,568	0.069	14,177,079	55,684.210	14,171,994	329.877	14,177,079	32.008
2000_75_10	7,788,713	0.069	7,815,419	48,717.480	7,815,755	201.636	7,815,755	20.537
2000_100_1	37,911,125	0.069	37,929,562	57,195.970	37,929,909	270.140	37,929,909	21.622
2000_100_2	33,606,556	0.069	33,665,281	57,844.250	33,647,322	490.736	33,665,281	34.322
2000_100_3	29,876,545	0.069	29,951,509	57,198.420	29,952,019	923.360	29,952,019	23.249
2000_100_4	26,925,001	0.068	26,948,234	57,484.560	26,949,268	440.690	26,949,268	23.800
2000_100_5	22,009,071	0.070	22,040,523	58,316.780	22,041,715	466.252	22,041,715	23.346
2000_100_6	18,850,028	0.070	18,868,630	56,282.860	18,868,887	339.878	18,868,887	22.315
2000_100_7	15,829,907	0.070	15,850,198	54,333.570	15,850,597	358.472	15,850,597	22.555
2000_100_8	13,607,904	0.070	13,628,210	52,206.350	13,628,967	231.923	13,628,967	22.250
2000_100_9	8,374,905	0.071	8,394,440	45,817.310	8,394,562	188.672	8,394,562	18.686
2000_100_10	4,909,671	0.070	4,923,413	38,243.750	4,923,559	124.031	4,923,559	15.041

δ's vanish, then a mean field equilibrium distribution like (3) perfectly represents the statistical properties of the system. It's generally believed that this is the case in the limit of very large system sizes under suitable conditions for the exchange interactions J's. An excellent account of mean field methods applied to spin glasses for non-physicists can be found in Opper and Saad (2001), from which the following explanation is based. A general case for which vanishing fluctuations in the infinite size limit is expected is when the $J_{i,j}$'s are of *infinite range*. The phrase infinite range is best understood if

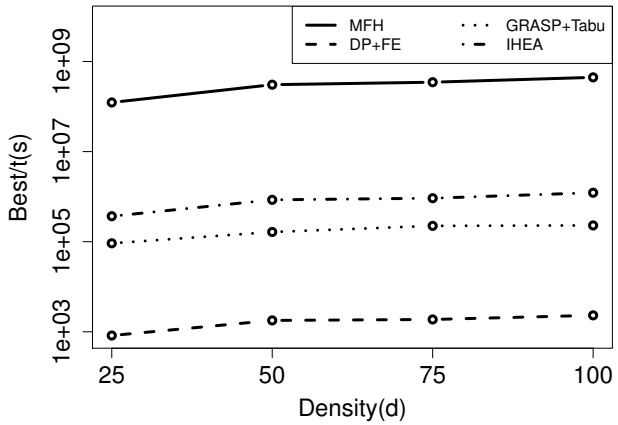

Figure 5 **Dual mean field annealing obtains the best performance as measured by the ratio between the best solution values and the corresponding computing times, when compared to state of the art heuristics for very large instances (1000 variables) of the Quadratic Knapsack Problem.**

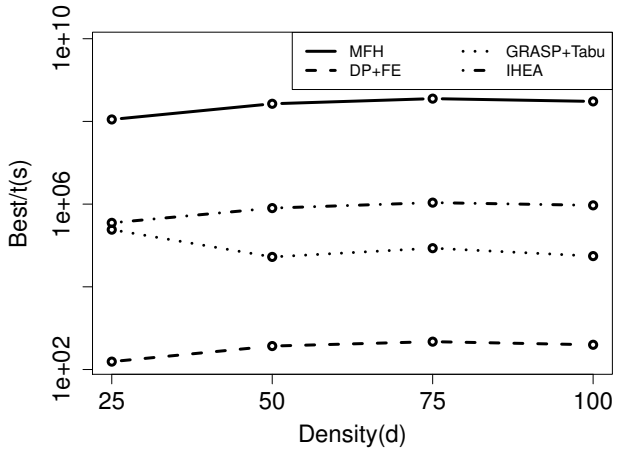

Figure 6 **Ratio between profit and time for different heuristics in cases of size N=2000 of QKP.**

we assume for a moment that the spins are located at sites i on a finite dimensional lattice. A spin glass model is said to be of infinite range if the $J_{i,j}$'s don't decay to zero when the distance $||i-j||$ is large. Note that when the connections $J_{i,j}$ between two arbitrary spins are random (including sparse or dense connectivities), the model is

Table 4 Comparative results of DMFA with 1 state-of-the-art algorithms on the 40 very large benchmark instances generated by Chen and Hao (2017).

N_d_l	DMFA		IHEA (Chen and Hao 2017)	
	Best	t(s)	Best	t(s)
5000_25_1	23,614,840	0.432	23,667,450	130.664
5000_25_2	37,833,091	0.421	37,914,560	143.679
5000_25_3	68,241,357	0.424	68,295,820	126.904
5000_25_4	33,811,271	0.431	33,866,053	139.453
5000_25_5	9,494,449	0.421	9,533,115	111.366
5000_50_1	45,173,276	0.420	45,194,685	144.125
5000_50_2	88,291,369	0.443	88,355,678	143.188
5000_50_3	152,355,756	0.426	152,447,303	143.813
5000_50_4	170,934,350	0.423	171,000,228	148.015
5000_50_5	1,181,814	0.424	1,187,339	61.106
5000_75_1	28,128,149	0.439	28,170,819	105.745
5000_75_2	195,365,500	0.426	195,434,758	149.977
5000_75_3	64,272,514	0.430	64,324,704	141.571
5000_75_4	247,307,461	0.420	247,348,595	144.213
5000_75_5	46,403,603	0.418	46,462,750	136.119
5000_100_1	214,357,297	0.413	214,425,886	150.076
5000_100_2	18,758,392	0.427	18,783,132	76.661
5000_100_3	10,768,973	0.425	10,784,650	61.45
5000_100_4	160,509,127	0.426	160,539,947	153.082
5000_100_5	33,135,895	0.412	33,166,524	105.708
6000_25_1	69,789,341	0.609	69,832,542	204.23
6000_25_2	3,673,316	0.609	3,697,236	123.77
6000_25_3	79,246,984	0.605	79,300,092	246.285
6000_25_4	191,462,861	0.591	191,531,304	238.917
6000_25_5	36,080,566	0.614	36,121,510	208.762
6000_50_1	194,286,675	0.591	194,344,567	214.187
6000_50_2	323,635,258	0.609	323,753,804	272.235
6000_50_3	31,859,452	0.615	31,913,824	220.343
6000_50_4	225,486,858	0.598	225,556,641	198.893
6000_50_5	40,885,584	0.603	40,931,924	186.351
6000_75_1	204,401,513	0.599	204,512,250	267.433
6000_75_2	42,353,252	0.604	42,422,207	182.99
6000_75_3	524,314,637	0.604	524,508,156	177.873
6000_75_4	196,912,744	0.598	197,004,931	220.513
6000_75_5	74,332,278	0.610	74,350,712	282.668
6000_100_1	292,228,015	0.610	292,257,056	219.599
6000_100_2	219,704,463	0.610	219,791,358	257.679
6000_100_3	376,854,314	0.587	376,967,122	266.202
6000_100_4	355,591,433	0.600	355,609,720	245.857
6000_100_5	686,256,995	0.612	686,364,195	211.295

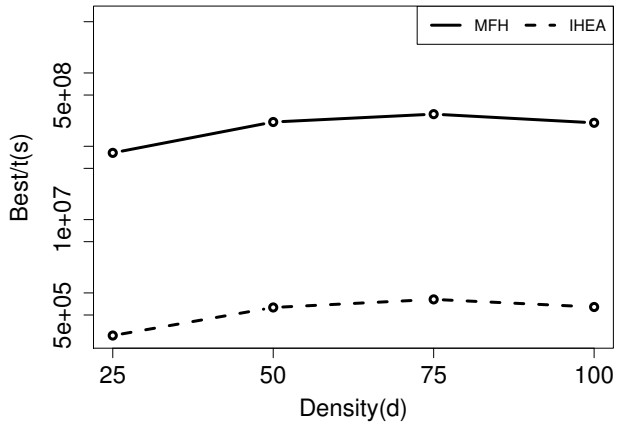

Figure 7 **Ratio between profit and time for different heuristics in cases of size N=5000 of QKP.**

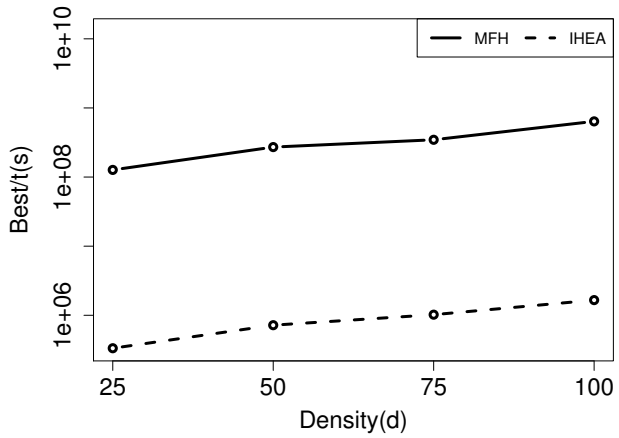

Figure 8 **Ratio between profit and time for different heuristics in cases of size N=6000 of QKP.**

trivially of infinite range. This of course is the situation for the objective function of the considered QKP instances, in which there is no reference to an underlying lattice and the values for the $q_{i,j}$'s are random. The crucial difference between QKP and infinite range spin glasses obviously is the linear constraint. Our experimental results however indicate a similar behavior of the mean field approximation in both models. This in our opinion opens a valuable line of research that could have significant consequences for

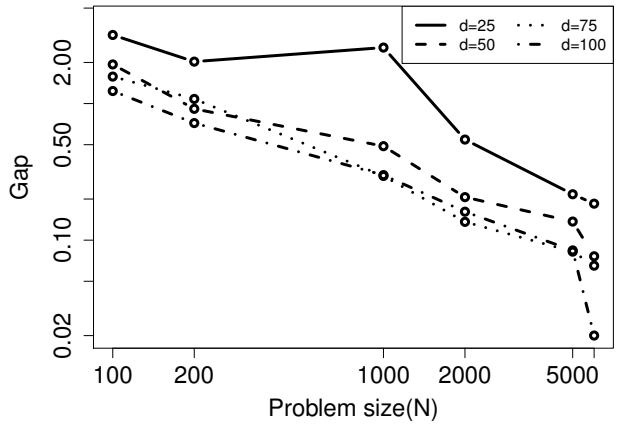

Figure 9 **Gaps of the dual mean field solutions with respect to the best solutions known for all of the considered QKP instances.**

computational complexity issues and heuristics construction in the context of nonlinear binary programming.

7. Conclusions

A principled heuristic for binary programming has been presented and its effectivity has been tested on linear and quadratic knapsack problems, for which it displays competitive performance. Besides its potential for problem solving, the approach suggest links between binary optimization and the theory of spin glasses that might result particularly relevant for large scale nonlinear binary optimization.

Acknowledgments

This work was partially supported by the National Council of Science and Technology of Mexico under grant CONACYT CB-167651, Cátedras-CONACYT no. 2193-825 and by the Autonomous University of Nuevo Leon support to research program under grant UANL-PAICYT IT451-15.

References

Banitalebi A, Aziz MIA, Aziz ZA (2016) A self-adaptive binary differential evolution algorithm for large scale binary optimization problems. *Information Sciences* 367–368(Nov):487–511.

Billionnet A, Soutif E (2004) Using a Mixed binary Programming Tool for Solving the 0-1 Quadratic Knapsack Problem. *INFORMS Journal on Computing* 16(2):188-197.

Billionnet A, Soutif E (2004) An exact method based on Lagrangian decomposition for the 0–1 quadratic knapsack problem. *Eur J Oper Res* 157(3):565–575.

Boyer V, El Baz D, Elkihel M (2012) Solving knapsack problems on GPU. *Computers & Operations Research* 39(1):42-47.

Chen Y, Hao JK (2017) An iterated "hyperplane exploration" approach for the quadratic knapsack problem. *Computers & Operations Research* 77:226–239.

Chhajed D, Lowe TJ (2008) *Building intuition: insights from basic operations management models and principles* (Vol. 115). (Springer Science & Business Media).

Chong, EKP, Zak SH (2013) *An Introduction to Optimization*, fourth ed. (Wiley).

Dolgui A, Kovalev S, Pesch E (2015) Approximate solution of a profit maximization constrained virtual business planning problem. *Omega* 57:212–216.

Fomeni FD, Letchford AN (2013) A dynamic programming heuristic for the quadratic knapsack problem. *INFORMS Journal on Computing* 26(1):173–182.

Hogg T (1996) Refining the phase transition in combinatorial search. *Artificial Intelligence* 81(1):127–154.

Jaynes ET (1957) Information theory and statistical mechanics. *Physical Review* 106(4):620.

Karimi GR, AziziVerki A (2012) *Mean Field Annealing Based Techniques for Resolving VLSI Automatic Design Problems*. In Simulated Annealing - Single and Multiple Objective Problems, Dr. Marcos Sales Guerra Tsuzuki (Ed.), Intech.

Kellerer H, Pferschy U, Pisinger D (2004) *Introduction to NP-Completeness of knapsack problems* (Springer Berlin Heidelberg).

Kirkpatrick S, Gelatt Jr. CD, Vecchi, MP (1983) Optimization by Simulated Annealing, *Science* 220(4598):671–680.

Kong X, Gao L, Ouyang H, Li S (2015) A simplified binary harmony search algorithm for large scale 0–1 knapsack problems. *Expert Systems with Applications* 42(12):5337–5355.

Luss H (2012) *Equitable Resource Allocation: Models, Algorithms and Applications* (Vol. 101). (John Wiley & Sons).

Martin OC, Monasson R, Zecchina R (2001) Statistical mechanics methods and phase transitions in optimization problems. *Theoretical Computer Science* 265(1):3–67.

Metropolis N, Rosenbluth A, Rosenbluth M, Teller A, Teller E (1953) Equation of state calculations by fast computing machines. *J. of Chemical Physics* 21(6):1087–1092.

Ohlsson M, Pi H (1995) A study of the mean field approach to knapsack problems. *Neural Networks* 10(2):263–271.

Ohlsson M, Peterson C, Söderberg B (2001) An efficient mean field approach to the set covering problem. *European Journal of Operational Research* 133(3):583–595.

Opper M, Saad D (2001) *Advanced Mean Field Methods: Theory and Practice* (MIT Press).

Parisi G (1988) *Statistical Field Theory* (Addison-Wesley).

Pisinger D (2005) Where are the hard knapsack problems?. *Computers and Operations Research* 32(9):2271–2284.

Pisinger D (2007) The quadratic knapsack problem–a survey. *Discrete Applied Mathematics* 155(5):623–648.

Roland J, Figueira JR, De Smet Y (2016) Finding compromise solutions in project portfolio selection with multiple experts by inverse optimization. *Computers & Operations Research* 66:12–19.

Sherrington D, Kirkpatrick S (1975) Solvable model of a spin-glass. *Physical review letters* 35(26):1792.

Thouless DJ, Anderson PW, Palmer RG (1977) Solution of "solvable model of a spin glass". *Philosophical Magazine* 35(3):593-601.

Yang Z, Wang G, Chu F (2013) An effective GRASP and tabu search for the 0–1 quadratic knapsack problem. *Computers & Operations Research* 40(5):1176–1185.

Young AP (1997) *Spin glasses and random fields (Vol. 12)* (World Scientific).

Yuan X, Nie H, Su A, Wang L, Yuan Y (2009) An improved binary particle swarm optimization for unit commitment problem. *Expert Systems with applications* 36(4):8049–8055.