Artificial Intelligence 175 (2011) 914–927

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

On the phase transitions of random k-constraint satisfaction problems

Yun Fan [a], Jing Shen [a,b,*]

[a] Department of Mathematics, Central China Normal University, Wuhan, 430079, China
[b] School of Science, Naval University of Engineering, Wuhan, 430033, China

ARTICLE INFO

Article history:
Received 29 July 2010
Received in revised form 11 November 2010
Accepted 11 November 2010
Available online 13 November 2010

Keywords:
Constraint satisfaction problem
Phase transition

ABSTRACT

Constraint satisfaction has received increasing attention over the years. Intense research has focused on solving all kinds of constraint satisfaction problems (CSPs). In this paper, first we propose a random CSP model, named k-CSP, that guarantees the existence of phase transitions under certain circumstances. The exact location of the phase transition is quantified and experimental results are provided to illustrate the performance of the proposed model. Second, we revise the model k-CSP to a random linear CSP by incorporating certain linear structure to constraint relations. We also prove the existence of the phase transition and exhibit its exact location for this random linear CSP model.

1. Introduction

Constraint satisfaction problem, or CSP in short, is represented by a finite set of variables, each one of which is associated with a domain, and a finite set of *constraints*, each of which consists of a subset of the variables, called a *constraint scope*, and a *constraint relation* that restricts the values of the variables in the constraint scope can simultaneously take. The objective is to assign a value to each variable satisfying all the constraint relations. CSP is an important topic in the area of computer science, especially in artificial intelligence, since the regularity in their formulation provides a common base to analyze and solve the problems of many unrelated families. In recent years, the random CSP and the corresponding phase transitions have attracted more and more attention since Cheeseman et al. proposed in [7] that many hard instances should be found at the phase transition points.

There have been various models for investigating the phase transitions of random CSP proposed by various academic communities, e.g. [2,10–12,17–19,27–29]. The initial standard models, named A, B, C and D [23,28], were proposed to generate random binary CSP instances. Experiments showed that the standard models [23,28] all exhibit a "threshold-like" behavior. On the other hand, it has been proved theoretically by Achlioptas et al. in [1] that the random instances generated by the standard models do not have an asymptotic threshold when the length of constraint scopes and the size of domains are fixed.

Improvement of the performance of standard models was addressed from various perspectives in numerous efforts [1,21–23,26,31]. Some new models incorporated special combinatorial structures on the constraints. In other words, the constraints are subject to certain combinatorial restrictions and the restrictions ensure that the generated instances are arc consistent [23], path consistent [21], strongly 3-consistent [22] or weakly 4-consistent [22]. It has been proved that all these revised models have non-trivial asymptotic behaviors. While the combinatorial structures provide the capability for producing phase transitions, this achievement typically comes at the price of more restrictions on the constraint relations of instances.

Based on the model B [28] mentioned previously, Xu and Li [31] proposed a random CSP model, named *model RB*. Instead of fixing the size of domains associated with each instance as in the model B, the size of domains of the model RB

* Corresponding author at: Department of Mathematics, Central China Normal University, Wuhan, 430079, China.
E-mail addresses: yunfan02@yahoo.com.cn (Y. Fan), shendina@hotmail.com (J. Shen).

0004-3702/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.artint.2010.11.004

is uniform for each instance and the value of the size is variant as a power function of the cardinality of the set of variables, i.e. the number of variables. On the other hand, the length of constraint scopes and the tightness of constraint relations for each instance of the model RB are fixed. It has been proved theoretically in [1] that model B does not have the phase transition point over most of the parameter space. Due to incorporating uniformly variant domain size, the revised model RB does have phase transitions and the exact phase transition points have been quantified by Xu and Li [31]. Moreover, it has been demonstrated that the model RB has a lot of hard instances existing [32] and all the instances at the phase transition points have exponential tree-resolution complexity [30]. Compared with revised models in [21–23], the capability of generating instances is dramatically enhanced for the model RB due to the fact that there is no combinatorial restriction enforced on the constraint relations of model RB. Generally speaking, it is natural and relatively easier for the model RB to generate asymptotically non-trivial CSP instances with relatively large domain size.

Another area of active research in the field of CSP is the development of k-SAT, where k denotes the length of the constraint scopes. It is proved by Friedgut in [17] that a phase transition exists for k-SAT if k is fixed. However, for fixed k with $k \geqslant 3$, there is still no effective method to obtain the exact location of the phase transition. For example, it is already derived theoretically in [24] and [14] that the best lower bound and upper bound of the phase transition for 3-SAT are 3.53 and 4.506 respectively; but the exact location is still under investigation. When compared with the results for k-SAT with fixed k, it has been demonstrated that it is possible for k-SAT to ascertain the exact location of phase transitions if the parameter k is growing moderately, as detailed in [16,20].

Motivated by k-SAT with growing k in [20], in this paper first we revise the model RB in [31] to propose a new random CSP model, named *model k-CSP*. When comparing with the model RB in [20], instead of fixing the parameters of constraints, including both the length of constraint scopes and the tightness of constraint relations, and varying the size of domains as a power function, for the new model k-CSP we assume that the size of domains and the tightness of constraint relations are fixed and the length of constraint scopes, which is denoted by k, is variant as a function of n, where n denotes the cardinality of the set of variables. Although similar to k-SAT, the new proposed model k-CSP has growing length of constraint scopes, the two models are essentially different from each other, more specifically, the tightness of constraint relations is variant for k-SAT while is fixed for the model k-CSP. For the new model k-CSP we theoretically prove the existence of a phase transition when the parameter k grows up to a logarithm function of n, and determine the exact location of the phase transition point. Further, we experimentally demonstrate the performance of the proposed model k-CSP. The experiments we conducted on the k-CSP not only verify the theoretical results we established, but also illustrate that the computational complexity of the k-CSP grows exponentially with n (the number of variables) and the worse-cases happen around the phase transition point. We note that, the model k-CSP can generate instances as easily as the model RB since there is no other restrictions on the constraint relations except the fixed tightness; on the other hand, the parameter k of the model k-CSP is growing up very slowly as a logarithm function rather than the power function appearing in the model RB. In summary, the model k-CSP can easily and naturally generate asymptotically non-trivial CSP instances within a reasonably small range of domain size and constraint scope length, thus it is very suitable for testing the capability of CSP algorithms.

The algebraic CSP, which employs algebraic structures to the domains and the constraint relations of CSP model, is another popular approach to construct a CSP model. We note that the algebraic CSP approach has received considerable attention in recent years [4]. One classical example of algebraic CSPs is the *linear CSP*, which domains are finite fields and constraint relations are affine subspaces of the vector spaces over the finite fields. One of the major advantages of various linear CSP models [3,5,6,8,9,13,15] is that they all exhibit satisfiability thresholds. This motivates our other research in constructing a random CSP model that combines the model k-CSP mentioned above with the linear CSP model.

To combine the advantages of linear CSP and the model k-CSP, we incorporate certain algebraic structure to the domains and constraint relations of k-CSP and then introduce another type of random linear CSP model, named *k-hyper-$\mathbb{F}$-linear CSP*. For each instance of the new proposed model, we assume that the domain could be any finite field, which is denoted by $\mathbb{F}$; the constraint relations are randomly chosen from the hyperplanes of the vector space $\mathbb{F}^k$, where k is the length of constraint scopes. Similar to the model k-CSP, the length of constraint scopes k is uniformly variant as a function of n, where n denotes the number of variables. We exhibit theoretically the exact phase transition of the model k-hyper-$\mathbb{F}$-linear CSP. When comparing with the linear CSP models from [3,5,6,8,9,13], we provide a more general formulation and a new proof based on a more general argument, which make the k-hyper-$\mathbb{F}$-linear CSP model more widely applicable in practice.

This paper is organized as follows. Section 2 states some preliminary definitions, introduces the random model k-CSP and presents the main theorem on the exact phase transition of the model k-CSP. Section 3 provides the complete proof of the theorem stated in Section 2. Section 4 summarizes our experimental results and analyzes the performance of the model k-CSP. Section 5 proposes the model k-hyper-$\mathbb{F}$-linear CSP and quantifies the exact phase transitions of the proposed model. Conclusions are provided in Section 6.

2. Random model *k*-CSP

In this paper, $\ln x = \log_e x$ denotes the natural logarithm function where e denotes the natural base, $\exp x = e^x$ denotes the natural exponential function, and $H(x) = -x\ln x - (1-x)\ln(1-x)$ for $x \in [0,1]$ denotes the natural entropy function. By $|T|$ for any set T we denote the cardinality of the set T.

A *constraint satisfaction problem*, CSP in short, is described as follows:

Instance. A triple $I = (X, D, C)$ where

- $X = (x_1, \ldots, x_n)$ is a sequence of n *variables*;
- $D = (D_1, \ldots, D_n)$ is a sequence of finite sets, called the *domains* of the instance;
- $C = (C_1, \ldots, C_t)$ with each $C_i = (X_i, R_i)$, called *constraints*, such that
  - $*$ $X_i = (x_{i_1}, \ldots, x_{i_{k_i}})$ is a subsequence of X of length k_i, called the *constraint scopes*,
  - $*$ R_i is a subset of $D_{i_1} \times \cdots \times D_{i_{k_i}}$, called the *constraint relations*.

Question. Is there a map f from X to the disjoint union $\bigcup_{i=1}^{n} D_i$ with each $f(x_i) \in D_i$ satisfying that $f(X_i) := (f(x_{i_1}), \ldots, f(x_{i_{k_i}})) \in R_i$ for all $i = 1, \ldots, t$?

If such a map f exists, then we would say that the instance I is *satisfiable* and $f(X) = (f(x_1), \ldots, f(x_n))$ is a *solution* of the instance I.

Let $A = D_1 \times \cdots \times D_n$. Any map $f(X) = (f(x_1), \ldots, f(x_n)) \in A$, i.e. any element $(a_1, \ldots, a_n) \in A$, is also said to be an *assignment* to the variables X with values in A.

It is interesting to randomize CSP and consider the asymptotic property of the random CSP.

Revising the model RB from [31] and k-SAT with growing k, we introduce a random model of CSP as follows.

Definition 2.1. A random CSP model is said to be *k-CSP* if the instances are generated as follows:

- every cardinality $|D_i| = d$ for $i = 1, \ldots, n$, where $d > 1$ is the size of domains;
- $t = t(n)$ is an integer function of n such that $\lim_{n\to\infty} t(n) = \infty$;
- for $i = 1, \ldots, t$ the constraints are generated as follows:
  - $*$ the constraint scopes $X_i = (x_{i_1}, \ldots, x_{i_k})$ with length $k = k(n)$, which is an integer function of n, are randomly selected with repetition allowed;
  - $*$ the constraint relations R_i are randomly selected with repetition allowed from the subsets of $D_{i_1} \times D_{i_2} \times \cdots \times D_{i_k}$ such that the cardinality $|R_i| = pd^k$, where p represents the tightness of the constraint relations and is a real constant with $0 < p < 1$.

By Pr(SAT) we denote the probability of a random instance of the model k-CSP being satisfiable. We have the following asymptotic properties of the model k-CSP.

Theorem 2.1. *Keep the same notation as in Definition* 2.1, *and assume that* $t = r \cdot \frac{n \ln d}{-\ln p}$ *for a constant parameter* r. *If* $k(n) \geqslant (2+\varepsilon)\frac{\ln n}{\ln d}$ *for a real* $\varepsilon > 0$, *then*

$$\lim_{n\to\infty} \Pr(\mathrm{SAT}) = \begin{cases} 0, & r > 1; \\ 1, & r < 1. \end{cases}$$

3. Proof of Theorem 2.1

Given any n, let $\mathcal{G}$ denote the set of all the instances of the model k-CSP with $X = (x_1, \ldots, x_n)$ and $D = (D_1, \ldots, D_n)$. Then $\mathcal{G}$ is a probability space with equal probability for all samples. For $I \in \mathcal{G}$, let $\mathrm{Sol}(I)$ denote the set of solutions of the instance I.

For any $\mathbf{a} = (a_1, \ldots, a_n) \in A$, let

$$S_{\mathbf{a}} = \begin{cases} 1, & \text{if } \mathbf{a} \in \mathrm{Sol}(I); \\ 0, & \text{otherwise.} \end{cases}$$

Then $S_{\mathbf{a}}$ is a 0–1 random variable over the probability space $\mathcal{G}$. And

$$S = \sum_{\mathbf{a}\in A} S_{\mathbf{a}} \tag{1}$$

is a non-negative integer random variable over the probability space $\mathcal{G}$.

The random variable S is the number of solutions of the random instance $I \in \mathcal{G}$; in particular, the probability $\Pr(S > 0)$ is just the probability Pr(SAT) for the random instance I being satisfiable.

Assume that $\mathbf{a} \in A$. Then $\Pr(\mathbf{a} \in R_i) = p$ as $|R_i| = pd^k$. Since $R_1, \ldots, R_t$ are selected randomly independently, we have

$$\Pr(S_{\mathbf{a}} = 1) = \prod_{i=1}^{t} \Pr(\mathbf{a} \in R_i) = p^t. \tag{2}$$

So the expectation of S is given by

$$\mathrm{E}(S) = \sum_{\mathbf{a}\in A} \mathrm{E}(S_{\mathbf{a}}) = d^n p^t. \tag{3}$$

By the assumption of Theorem 2.1 that $t = r \cdot \frac{n \ln d}{-\ln p}$, we have

$$d^n p^t = \exp(n \ln d + t \ln p) = \exp\bigl(n(1-r)\ln d\bigr) = d^{(1-r)n}. \tag{4}$$

Hence $\lim_{n\to\infty} d^n p^t = 0$ if $r > 1$; and by Markov's inequality

$$\Pr(S > 0) \leqslant \mathrm{E}(S) = d^n p^t,$$

we have

$$\lim_{n\to\infty} \Pr(S > 0) = 0, \quad \text{if } r > 1.$$

In the rest of this section we always assume that

$$t = \frac{rn \ln d}{-\ln p}, \quad 0 < r < 1, \tag{5}$$

and aim at proving, with the conditions of Theorem 2.1, that

$$\lim_{n\to\infty} \Pr(S > 0) = 1. \tag{6}$$

By the convexity of the function $\frac{1}{x}$ for $x > 0$, it follows from the Jensen's inequality that

$$\Pr(S > 0) \geqslant \sum_{\mathbf{a}\in A} \frac{\Pr(S_{\mathbf{a}} = 1)}{\mathrm{E}(S|S_{\mathbf{a}} = 1)}, \tag{7}$$

where $\mathrm{E}(S|S_{\mathbf{a}} = 1)$ denotes the conditional expectation of S assuming that $S_{\mathbf{a}} = 1$ occurs; see [25, Theorem 6.10]. By the way, we remark that the following argument based on this inequality is in fact equivalent to the so-called second moment method, please see Appendix A of the paper for details.

Let $\mathbf{a} = (a_1, \ldots, a_n) \in A$ and $\mathbf{b} = (b_1, \ldots, b_n) \in A$ with $\mathbf{a} \neq \mathbf{b}$. We calculate the probability of both $\mathbf{a}$ and $\mathbf{b}$ satisfying a random instance $I \in \mathcal{G}$, where I has constraints $X_i = (x_{i_1}, \ldots, x_{i_k})$ and $R_i \subseteq D_{i_1} \times \cdots \times D_{i_k}$ for $i = 1, \ldots, t$.

Let m be the number of such indices i that $a_i = b_i$, i.e. the defect m of the Hamming distance between $\mathbf{b}$ and $\mathbf{a}$. There are two cases:

$*$ either, $\mathbf{a}$ and $\mathbf{b}$ agree with each other on every variable of the constraint, in this case, the conditional probability of both $\mathbf{a}$ and $\mathbf{b}$ satisfying the constraint relation R_i is

$$\binom{d^k - 1}{pd^k - 1} \Big/ \binom{d^k}{pd^k} = p;$$

$*$ or, the conditional probability of $\mathbf{a}$ and $\mathbf{b}$ satisfying the constraint relation R_i is

$$\binom{d^k - 2}{pd^k - 2} \Big/ \binom{d^k}{pd^k} = p\left(\frac{pd^k - 1}{d^k - 1}\right).$$

The probability that the first case occurs is $\sigma_{m,n} = \frac{\binom{m}{k}}{\binom{n}{k}}$. Thus we obtain that

$$\Pr(\mathbf{a} \in R_i, \mathbf{b} \in R_i) = p \cdot \sigma_{m,n} + p\left(\frac{pd^k - 1}{d^k - 1}\right)(1 - \sigma_{m,n}).$$

Since the constraint relations $R_1, \ldots, R_t$ are selected independently, we get that

$$\begin{aligned}\Pr(S_{\mathbf{a}} = 1, S_{\mathbf{b}} = 1) &= \prod_{i=1}^{t} \Pr(\mathbf{a} \in R_i, \mathbf{b} \in R_i) \\ &= p^t \left(\sigma_{m,n} + \frac{pd^k - 1}{d^k - 1}(1 - \sigma_{m,n})\right)^t.\end{aligned}$$

According to conditional probability, we get

$$\Pr(S_{\mathbf{b}} = 1 | S_{\mathbf{a}} = 1) = \left(\sigma_{m,n} + \frac{pd^k - 1}{d^k - 1}(1 - \sigma_{m,n})\right)^t.$$

It is clear that $\mathrm{E}(S_{\mathbf{b}}|S_{\mathbf{a}}=1)=\Pr(S_{\mathbf{b}}=1|S_{\mathbf{a}}=1)$ and $\mathrm{E}(S|S_{\mathbf{a}}=1)=\sum_{\mathbf{b}\in A}\mathrm{E}(S_{\mathbf{b}}|S_{\mathbf{a}}=1)$. In conclusion, we have

$$\mathrm{E}(S|S_{\mathbf{a}}=1)=\sum_{m=0}^{n}\binom{n}{m}(d-1)^{n-m}\left(\sigma_{m,n}+(1-\sigma_{m,n})\frac{pd^k-1}{d^k-1}\right)^t. \tag{8}$$

Further

$$\frac{pd^k-1}{d^k-1}=\frac{pd^k-p+p-1}{d^k-1}=p-\frac{1-p}{d^k-1}\leqslant p,$$

consequently we obtain the following inequality

$$\mathrm{E}(S|S_{\mathbf{a}}=1)\leqslant\sum_{m=0}^{n}\binom{n}{m}(d-1)^{n-m}\big(\sigma_{m,n}+(1-\sigma_{m,n})p\big)^t.$$

Combining it with the formulas (3) and (7), we deduce that

$$\begin{aligned}\Pr(S>0)&\geqslant\frac{d^np^t}{\sum_{m=0}^{n}\binom{n}{m}(d-1)^{n-m}(\sigma_{m,n}+(1-\sigma_{m,n})p)^t}\\&=\frac{d^np^t}{\sum_{m=0}^{n}\binom{n}{m}(d-1)^{n-m}(p+(1-p)\sigma_{m,n})^t}.\end{aligned}$$

So

$$\frac{1}{\Pr(S>0)}\leqslant\frac{\sum_{m=0}^{n}\binom{n}{m}(d-1)^{n-m}(p+(1-p)\sigma_{m,n})^t}{d^np^t}.$$

For $m=0,1,\dots,k-1$ we have $\sigma_{m,n}=\frac{\binom{m}{k}}{\binom{n}{k}}=0$, i.e.

$$\sum_{m=0}^{k-1}\binom{n}{m}(d-1)^{n-m}\big(p+(1-p)\sigma_{m,n}\big)^t=\sum_{m=0}^{k-1}\binom{n}{m}(d-1)^{n-m}p^t;$$

noting that $\sum_{m=0}^{n}\binom{n}{m}(d-1)^{n-m}=d^n$, we have

$$\frac{1}{\Pr(S>0)}\leqslant 1+\sum_{m=k}^{n}\frac{\binom{n}{m}(d-1)^{n-m}((p+(1-p)\sigma_{m,n})^t-p^t)}{d^np^t}.$$

For $m\geqslant k$, since $n\geqslant m$, we have $\frac{m-i}{n-i}<\frac{m}{n}$, hence $\sigma_{m,n}\leqslant\left(\frac{m}{n}\right)^k$; thus

$$\frac{1}{\Pr(S>0)}\leqslant 1+\sum_{m=k}^{n}\frac{\binom{n}{m}(d-1)^{n-m}((p+(1-p)(\frac{m}{n})^k)^t-p^t)}{d^np^t}. \tag{9}$$

For m with $k\leqslant m\leqslant n$ let

$$\begin{aligned}R_m&=\binom{n}{m}(d-1)^{n-m}\left(\left(p+(1-p)\left(\frac{m}{n}\right)^k\right)^t-p^t\right)\Big/d^np^t\\&=\binom{n}{m}\left(\frac{1}{d}\right)^m\left(1-\frac{1}{d}\right)^{n-m}\left(\left(1+\frac{1-p}{p}\left(\frac{m}{n}\right)^k\right)^t-1\right).\end{aligned}$$

In order to prove the equality (6), it is enough to show that, with the conditions of Theorem 2.1 and the assumption (5), we have

$$\lim_{n\to\infty}\sum_{m=k}^{n}R_m=0;$$

i.e. for any $\delta>0$ there is an integer N such that

$$\sum_{m=k}^{n}R_m<\delta,\quad\forall n>N.$$

Further, to prove the above, it is enough to show that for any m with $k\leqslant m\leqslant n$ we have

$$nR_m<\delta,\quad\forall n>N;$$

because this inequality implies that $\sum_{m=k}^{n}R_m<\sum_{m=k}^{n}\delta/n<\delta$.

It is known that $\binom{n}{m} < \exp(nH(m/n))$. Setting $x = \frac{m}{n}$ hence $m = nx$, we have

$$nR_m < n\exp\bigl(nH(x)\bigr) \cdot (d-1)^{n(1-x)}\bigl(\bigl(p+(1-p)x^k\bigr)^t - p^t\bigr)/d^n p^t.$$

Define the functions $f_n(x)$ for $n = 1, 2, \ldots$ as follows

$$f_n(x) = n\exp\bigl(nH(x)\bigr)(d-1)^{n(1-x)}\bigl(\bigl(p+(1-p)x^k\bigr)^t - p^t\bigr)/d^n p^t, \tag{10}$$

i.e.

$$f_n(x) = n\exp\bigl(nH(x)\bigr)\left(\frac{1}{d}\right)^{nx}\left(1-\frac{1}{d}\right)^{n(1-x)}\left(\left(1+\frac{1-p}{p}x^k\right)^t - 1\right). \tag{11}$$

Then, the following proposition is enough to complete a proof of Theorem 2.1.

Proposition 3.1. *Assume that $k(n) \geqslant (2+\varepsilon)\frac{\ln n}{\ln d}$ for a real $\varepsilon > 0$ and that (5) holds. Then for any $\delta > 0$ there is an integer N such that*

$$f_n(x) < \delta, \quad \forall n > N,\ \forall x \in (0, 1].$$

Proof. We prove it in three steps.

Step 1: we look for a positive real $\zeta < 1$ and an integer N_1 such that

$$f_n(x) < \delta, \quad \forall n > N_1,\ \forall x \in [\zeta, 1].$$

From formula (4), definition (10) and the fact that $p + (1-p)x^k \leqslant 1$, we have

$$\begin{aligned} f_n(x) &\leqslant n\exp\bigl(nH(x)\bigr)(d-1)^{n(1-x)}\bigl(p+(1-p)x^k\bigr)^t/d^{(1-r)n} \\ &\leqslant n\exp\bigl(nH(x)\bigr)(d-1)^{n(1-x)}/d^{(1-r)n} \\ &= \exp\bigl(n\bigl(n^{-1}\ln n + H(x) + (1-x)\ln(d-1) - (1-r)\ln d\bigr)\bigr). \end{aligned}$$

Denote $\tau = (1-r)\ln d$, which is a positive constant as $r < 1$. Since $H(x)$ is a continuous function with non-negative value on $[0, 1]$ and $H(1) = 0$, there is a positive real $\zeta_1 < 1$ such that

$$H(x) < \tau/4, \quad \forall x \in [\zeta_1, 1].$$

Obviously, we can take a positive real $\zeta_2 < 1$ such that

$$(1-x)\ln(d-1) < \tau/4, \quad \forall x \in [\zeta_2, 1].$$

On the other hand, since $\lim_{n\to\infty} n^{-1}\ln n = 0$, there is an integer N_{11} such that

$$n^{-1}\ln n < \tau/4, \quad \forall n > N_{11}.$$

Now, let $\zeta = \max\{\zeta_1, \zeta_2\}$; then

$$n^{-1}\ln n + H(x) + (1-x)\ln(d-1) - (1-r)\ln d < -\tau/4, \quad \forall n > N_{11},\ \forall x \in [\zeta, 1].$$

Take an integer $N_{12} > -4\ln\delta/\tau$, then

$$\exp(-n\tau/4) < \exp(\ln\delta) = \delta, \quad \forall n > N_{12}.$$

At last, take $N_1 = \max\{N_{11}, N_{12}\}$; we have

$$f_n(x) < \delta, \quad \forall n > N_1,\ \forall x \in [\zeta, 1].$$

Step 2: we show that there is an η with $1 > \eta > 0$ and an integer N_2 such that

$$f_n(x) < \delta, \quad \forall n > N_2,\ \forall x \in \left[\frac{1}{d} - \eta, \frac{1}{d} + \eta\right].$$

For the purpose, we show an easy lemma on the entropy function.

Lemma 3.1. *Assume that $a \in (0, 1)$. Let $H(x, a) = H(x) + x\ln a + (1-x)\ln(1-a)$ for $x \in (0, 1)$. Then $H(x, a)$ is strictly increasing in $(0, a)$, while $H(x, a)$ is strictly decreasing in $(a, 1)$; in particular, $H(x, a) < 0$ if $x \neq a$.*

Proof. It is derived from the following calculation of derivatives:

$$\frac{dH(x,a)}{dx} = -\ln x + \ln(1-x) + \ln a - \ln(1-a);$$
$$\frac{dH(x,a)}{dx} = 0 \quad \Longleftrightarrow \quad x = a;$$
$$\frac{d^2H(x,a)}{dx^2} = -\frac{1}{x} - \frac{1}{1-x} < 0.$$

Revising the formula (11) with the notation in Lemma 3.1, we have

$$f_n(x) = \exp\bigl(nH(x,1/d)\bigr)\cdot n\bigl(\bigl(1+(1-p)x^k/p\bigr)^t - 1\bigr). \tag{12}$$

By Lemma 3.1, $\exp(nH(x,1/d)) \leqslant 1$; so

$$f_n(x) \leqslant n\bigl(\bigl(1+(1-p)x^k/p\bigr)^t - 1\bigr).$$

Let $y = x^k$, $g(y) = (1+(1-p)y/p)^t$, then the derivative

$$\frac{dg(y)}{dy} = \frac{1-p}{p}t\left(1+\frac{1-p}{p}y\right)^{t-1};$$

and by the Mean Value Theorem, there is a $\theta_y \in (0,y)$ such that

$$\left(1+\frac{1-p}{p}x^k\right)^t - 1 = g(y) - g(0) = \frac{1-p}{p}t\left(1+\frac{1-p}{p}\theta_y\right)^{t-1} y.$$

Since $\theta_y \in (0,y) = (0,x^k)$, there is $\theta_x \in (0,x)$ such that $\theta_x^k = \theta_y$. Thus

$$\left(1+\frac{1-p}{p}x^k\right)^t - 1 = \frac{1-p}{p}t\left(1+\frac{1-p}{p}\theta_x^k\right)^{t-1} x^k \leqslant \frac{1-p}{p}t\left(1+\frac{1-p}{p}x^k\right)^{t-1} x^k.$$

And we have that $f_n(x) \leqslant n\frac{1-p}{p}t(1+\frac{1-p}{p}x^k)^{t-1}x^k$, i.e.

$$f_n(x) \leqslant \frac{1-p}{p}\exp\left(\ln n + \ln t + (t-1)\ln\left(1+\frac{1-p}{p}x^k\right) + k\ln x\right).$$

But, $(t-1)\cdot\ln(1+\frac{1-p}{p}x^k) \leqslant t\cdot\frac{1-p}{p}x^k$, and $t = \frac{rn\ln d}{-\ln p}$ by (5), we obtain

$$f_n(x) \leqslant \frac{1-p}{p}\exp\left(\ln\frac{r}{-\ln p} + \ln\ln d + 2\ln n + \frac{r(1-p)\ln d}{-p\ln p}nx^k + k\ln x\right).$$

Since $k \geqslant (2+\varepsilon)\ln n/\ln d$, we have

$$n(1/d)^k = nd^{-k} \leqslant n\bigl(d^{-\ln n/\ln d}\bigr)^{2+\varepsilon} = n^{-(1+\varepsilon)};$$

thus there is an $\eta_1 > 0$, a $M > 0$ and an integer N_{21} such that

$$\ln\frac{r}{-\ln p} + \ln\ln d + \frac{r(1-p)\ln d}{-p\ln p}nx^k < M, \quad \forall n > N_{21},\ \forall x \in \left[\frac{1}{d}-\eta_1, \frac{1}{d}+\eta_1\right].$$

On the other hand, by $k \geqslant (2+\varepsilon)\ln n/\ln d$ again, we have

$$2\ln n + k\ln(1/d) = 2\ln n - k\ln d < 2\ln n - (2+\varepsilon)\ln n = -\varepsilon\ln n;$$

hence there is an $\eta_2 > 0$ and an integer N_{22} such that

$$2\ln n + k\ln x < -M + \ln\bigl(p\delta/(1-p)\bigr), \quad \forall n > N_{22},\ \forall x \in \left[\frac{1}{d}-\eta_2, \frac{1}{d}+\eta_2\right].$$

Take $\eta = \min\{\eta_1, \eta_2\}$ and $N_2 = \max\{N_{21}, N_{22}\}$, then

$$f_n(x) < \delta, \quad \forall n > N_2,\ \forall x \in \left[\frac{1}{d}-\eta, \frac{1}{d}+\eta\right].$$

Step 3: there is an integer N_3 such that

$$f_n(x) < \delta, \quad \forall n > N_3,\ \forall x \in \left(0, \frac{1}{d}-\eta\right) \cup \left(\frac{1}{d}+\eta, \zeta\right).$$

This time we further revise the expression of $f_n(x)$ in the formula (12) as follows:

$$\begin{aligned}
f_n(x) &= \exp\left(nH\left(x, \frac{1}{d}\right)\right) \cdot n\left(\left(1+(1-p)x^k/p\right)^t - 1\right) \\
&\leqslant \exp\left(nH\left(x, \frac{1}{d}\right)\right) \cdot n\left(1+\frac{1-p}{p}x^k\right)^t \\
&= \exp\left(nH\left(x, \frac{1}{d}\right) + \ln n + t\ln\left(1+\frac{1-p}{p}x^k\right)\right) \\
&= \exp\left(nH\left(x, \frac{1}{d}\right) + \ln n + \frac{rn\ln d\ln(1+\frac{1-p}{p}x^k)}{-\ln p}\right) \\
&= \exp\left(n\left(H\left(x, \frac{1}{d}\right) + \frac{\ln n}{n} + \frac{r\ln d\ln(1+\frac{1-p}{p}x^k)}{-\ln p}\right)\right);
\end{aligned}$$

that is,

$$f_n(x) \leqslant \exp\left(n\left(H\left(x, \frac{1}{d}\right) + \frac{\ln n}{n} + \frac{r\ln d\ln(1+\frac{1-p}{p}x^k)}{-\ln p}\right)\right).$$

Let $-\sigma = \max\{H(\frac{1}{d}-\eta, \frac{1}{d}), H(\frac{1}{d}+\eta, \frac{1}{d})\}$. By Lemma 3.1, $\sigma > 0$ and

$$H\left(x, \frac{1}{d}\right) < -\sigma, \quad \forall x \in \left(0, \frac{1}{d}-\eta\right) \cup \left(\frac{1}{d}+\eta, \zeta\right).$$

On the other hand, since $\lim_{n\to\infty} \frac{\ln n}{n} = 0$, we have an integer N_{31} such that

$$\frac{\ln n}{n} < \frac{\sigma}{3}, \quad \forall n > N_{31}.$$

And, since $k \to \infty$ when $n \to \infty$, and $x < \zeta < 1$, hence

$$\lim_{n\to\infty} \ln\left(1+\frac{1-p}{p}x^k\right) = 0;$$

thus we have an integer N_{32} such that

$$\frac{r\ln d\ln(1+\frac{1-p}{p}x^k)}{-\ln p} < \frac{\sigma}{3}, \quad \forall n > N_{32},\ \forall x \in \left(0, \frac{1}{d}-\eta\right) \cup \left(\frac{1}{d}+\eta, \zeta\right).$$

Take an integer $N_{33} > -\frac{3\ln\delta}{\sigma}$, we have

$$\exp\left(-\frac{\sigma}{3}n\right) < \exp(\ln\delta) = \delta, \quad \forall n > N_{33}.$$

Now, letting $N_3 = \max\{N_{31}, N_{32}, N_{33}\}$, we have the wanted conclusion:

$$f_n(x) < \delta, \quad \forall n > N_3,\ \forall x \in \left(0, \frac{1}{d}-\eta\right) \cup \left(\frac{1}{d}+\eta, \zeta\right).$$

Summarizing the three steps and setting $N = \max\{N_1, N_2, N_3\}$, we reach the desired result of the proposition:

$$f_n(x) < \delta, \quad \forall n > N,\ \forall x \in (0, 1]. \qquad \square$$

4. Experimental results

In this section, we give some experimental results about the model k-CSP. The platform we have used for our experimentation is called Abscon (see http://www.cril.univ-artois.fr/~lecoutre). Note that Theorem 2.1 has guaranteed the existence of an asymptotical phase transition and located precisely the phase transition point.

Each random instance is characterized by a 5-tuple (k, n, d, r, p) of parameters, where k denotes the length of the constraint scopes, n the number of variables, d the uniform domain size, $r = \frac{-t\ln p}{n\ln d}$ a measure of the constraint density, p a measure of the constraint tightness. At each setting of (k, n, d, r, p), 50 instances are generated.

When the constraint density r is varied accordingly, the instances change from being soluble to insoluble. Fig. 1 depicts the solubility phase transition for $d = 4$, $p = 0.6$, $n \in \{20, 25, 30\}$ and $k = 5$, which is, by the condition of Theorem 2.1,

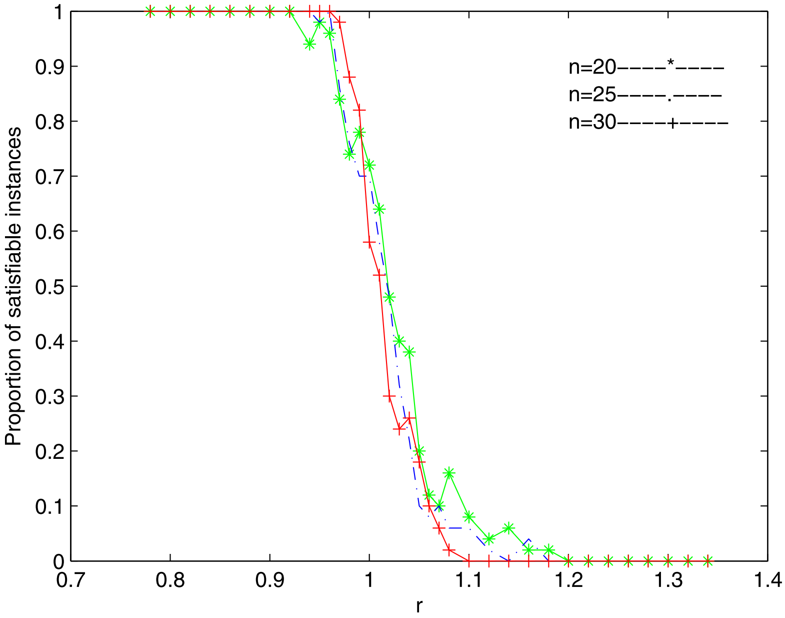

Fig. 1. The solubility phase transition for k-CSP $(5, \{20, 25, 30\}, 4, r, 0.6)$.

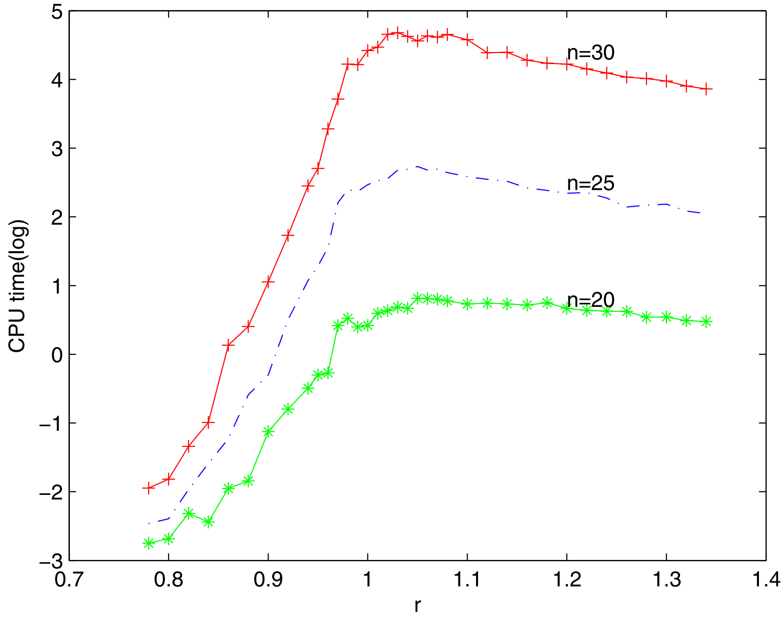

Fig. 2. Mean search cost of solving instances in k-CSP $(5, \{20, 25, 30\}, 4, r, 0.6)$.

the minimal value of k corresponding to $d = 4$ and $n \in \{20, 25, 30\}$. Note that the vertical scale of Fig. 1 refers to the proportion of satisfiable instances. Fig. 1 indicates that the experimental result supports affirmatively the theoretical result. Furthermore, it is interesting that, as shown in the picture, the model k-CSP exhibits the solubility phase transition even if the number n of variables is small, and the threshold interval becomes quickly narrow when the number n of variables increases.

Fig. 2 depicts the hardness of solving the instances of the model k-CSP when the constraint density r is varied. We select the values of parameters other than r for $k = 5$, $d = 4$, $p = 0.6$ and $n \in \{20, 25, 30\}$. In Fig. 2, it clearly appears that the hard instances are found at the neighborhood of the phase transition point $r = 1$. The solubility phase transition and the hardness phase transition both happen around the theoretical threshold $r = 1$ given by Theorem 2.1.

We have studied the computational complexity of solving the instances of the model k-CSP around the theoretical threshold $r = 1$ when n is varied from 10 to 50 in steps of 4. According to Theorem 2.1, k satisfies the condition $k \geqslant (2 + \varepsilon)\ln n / \ln d$

Table 1
The corresponding minimal value of k satisfying the condition against n.

n	10	14	18	22	26	30	34	38	42	46	50
k $(d=2)$	7	8	9	9	10	10	11	11	11	12	12
k $(d=5)$	3	4	4	4	5	5	5	5	5	5	5
k $(d=10)$	3	3	3	3	3	3	4	4	4	4	4

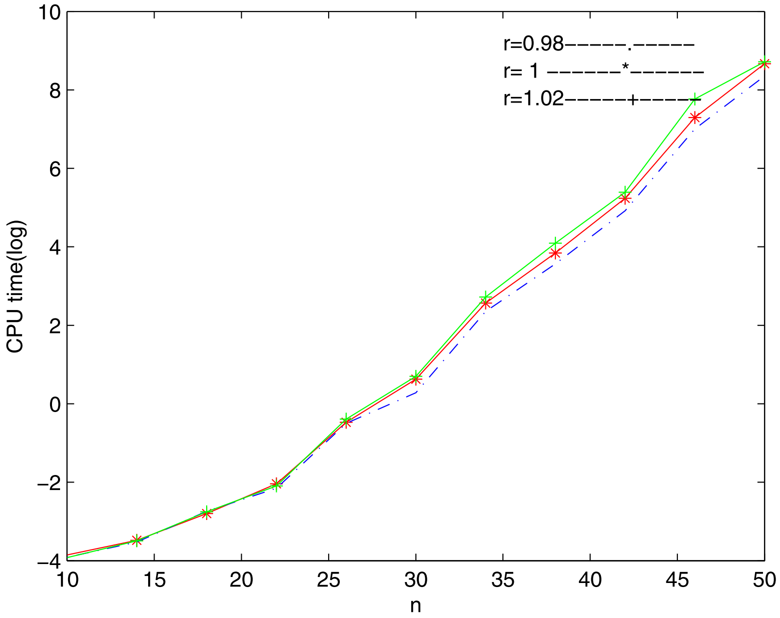

Fig. 3. Mean search cost of solving instances in k-CSP $(k, n, 2, r, 0.6)$.

for an arbitrary positive real ε. Table 1 gives the corresponding minimal value of k satisfying the condition against n for $d = 2$, $d = 5$, $d = 10$. The table shows that k increases very slowly with n. So it is valuable to generate random instances in experiments. We do experiments for $d = 2$, $p = 0.6$ and $r \in \{0.98, 1, 1.02\}$, and the result is shown in Fig. 3. Note that in Fig. 3 the horizontal scale uses a log scale. The curves in Fig. 3 look like the straight lines, which show that the complexity of solving the hard instances grows exponentially with n.

In addition, the effect of the domain size d and the constraint tightness p is investigated. We have studied the complexity of solving the hard instances at the theoretical threshold $r = 1$ according to the different values of d and p. Fig. 4 shows the computational complexity grows exponentially when d increases and the other parameters are fixed. Similarly, Fig. 5 indicates the computational complexity grows when p increases and the other parameters are fixed. The different values of d and p also illustrate the wide applicability of the model k-CSP.

5. A random model of linear CSP

In this section, we introduce a random model of linear CSP, which is corresponding to the model k-CSP in Section 3; and show a phase transition of it, which is corresponding to Theorem 2.1.

Let $\mathbb{F}$ be a finite field of order $q = \ell^a$ where ℓ is a prime integer and a is a positive integer. For any positive integer k, denote $\mathbb{F}^k = \mathbb{F} \times \cdots \times \mathbb{F}$ with k-multiple of $\mathbb{F}$. Then $\mathbb{F}^k$ is a vector space of dimension k over the field $\mathbb{F}$. Recall that, if U is a subspace of dimension d of the vector space $\mathbb{F}^k$ and $v \in \mathbb{F}^k$ is a vector, then the translation $v + U := \{v + u \mid u \in U\}$ of the subspace U is called an *affine subspace* of dimension d of $\mathbb{F}^k$. In that sense, $\mathbb{F}^k$ is also said to be an *affine space* of dimension k over the field $\mathbb{F}$. Note that a *hyperplane* of the affine space $\mathbb{F}^k$ means an affine subspace of $\mathbb{F}^k$ of dimension $k - 1$.

A CSP is said to be $\mathbb{F}$-*linear* if every instance $I = (X, D, C)$ satisfies further two conditions:

- the domains $D = (D_1, \ldots, D_n)$ are required such that $D_1 = \cdots = D_n = \mathbb{F}$;
- for every constraint scope $X_i = (x_{i_1}, \ldots, x_{i_{k_i}})$, $i = 1, \ldots, t$, the corresponding constraint relation R_i is an affine subspace of the affine space $\mathbb{F}^{k_i} = D_{i_1} \times \cdots \times D_{i_{k_i}}$.

It is again interesting to randomize the linear CSP and to consider the asymptotic property of the random linear CSP.

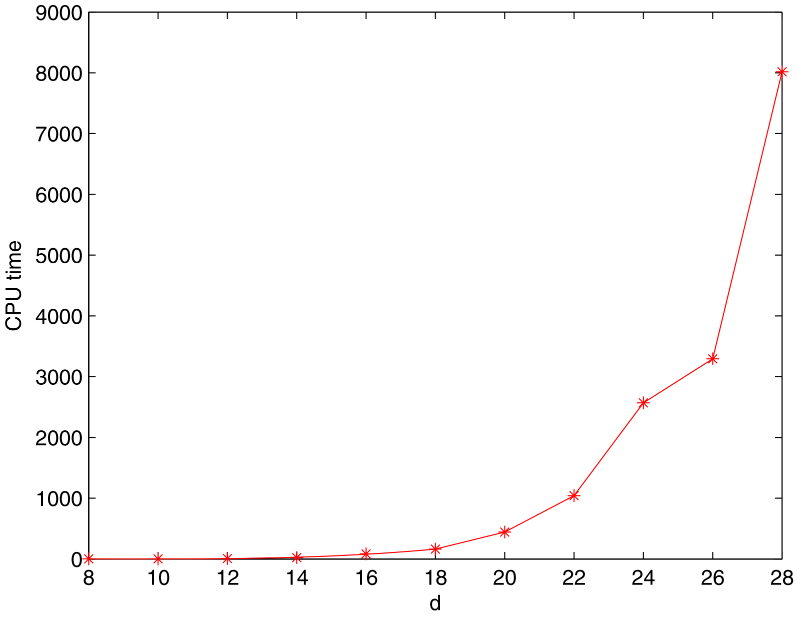

Fig. 4. Mean search cost of solving instances in k-CSP $(3, 20, d, 1, 0.6)$.

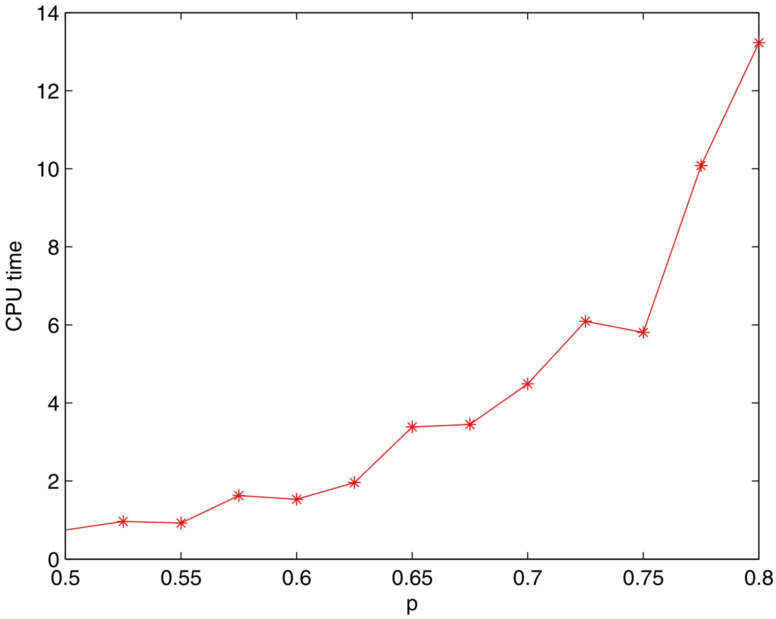

Fig. 5. Mean search cost of solving instances in k-CSP $(5, 20, 4, 1, p)$.

Definition 5.1. A random $\mathbb{F}$-linear CSP is said to be *k-hyper-$\mathbb{F}$-linear CSP* if the instances are generated as follows:

- $t = t(n)$ is an integer function of n such that $\lim_{n\to\infty} t(n) = \infty$;
- for $i = 1, \dots, t$ the constraints are generated as follows:
  - $*$ the constraint scopes $X_i = (x_{i_1}, \dots, x_{i_k})$ of length $k = k(n)$, which is an integer function of n, are randomly selected with repetition allowed;
  - $*$ the constraint relations R_i are randomly selected with repetition allowed from the set of the hyperplanes of $\mathbb{F}^k = D_{i_1} \times \cdots \times D_{i_k}$.

We denote Pr(SAT) again the probability of a random instance of the k-hyper $\mathbb{F}$-linear CSP being satisfiable, and have the following asymptotic properties.

Theorem 5.1. *Let notation be as in Definition* 5.1, *and assume that* $t = rn$ *for a constant parameter* r. *If* $k(n) \geqslant (2+\varepsilon)\frac{\ln n}{\ln q}$ *for a real* $\varepsilon > 0$, *then*

$$\lim_{n\to\infty} \Pr(\mathrm{SAT}) = \begin{cases} 0, & r > 1; \\ 1, & r < 1. \end{cases}$$

Remark. It is easy to compute the parameters:

- every domain D_i has constant cardinality $d = |D_i| = |\mathbb{F}| = q$;
- every R_i is a hyperplane of $\mathbb{F}^k$, hence $|R_i| = q^{k-1} = \frac{1}{q} \cdot d^k = pd^k$, where denote $p = \frac{1}{q}$;
- $t = rn = r\frac{n \ln d}{-\ln p}$, because $\frac{\ln d}{-\ln p} = \frac{\ln q}{-\ln q^{-1}} = 1$;
- $k(n) \geqslant (2+\varepsilon)\frac{\ln n}{\ln d}$, as $q = d$.

Thus all the conditions for the parameters in Theorem 2.1 are satisfied. Note that we cannot quote Theorem 2.1 to get Theorem 5.1 directly, because the probability space in the present case is different from that for Theorem 2.1.

Proof of Theorem 5.1. For any given n, let $\mathcal{L}$ be the probability space consisting of all the instances of $X = (x_1, \ldots, x_n)$ of the k-hyper-$\mathbb{F}$-linear CSP. Similarly to Section 3, we set $A = \mathbb{F}^n$, and define the random variable $S_{\mathbf{a}}$ for $\mathbf{a} \in A$ and $S = \sum_{\mathbf{a}\in A} S_{\mathbf{a}}$ as in the formula (1).

Further, each constraint relation R_i is a hyperplane of $\mathbb{F}^k$; the cardinality of the set of the hyperplanes of $\mathbb{F}^k$ is $\frac{(d^k-1)d}{d-1}$. It is clear that

$$\Pr(\mathbf{a} \in R_i) = \frac{d^k-1}{d-1} \Big/ \frac{(d^k-1)d}{d-1} = \frac{1}{d} = p.$$

We still obtain that

$$\Pr(S_{\mathbf{a}} = 1) = \Pr\bigl(\mathbf{a} \in \mathrm{Sol}(I)\bigr) = p^t;$$

and similarly to that in Section 3, the expectation $\mathrm{E}(S) = d^n p^t$.

For $\mathbf{a} \in A$, $\mathbf{b} \in A$, and $\mathbf{a} \neq \mathbf{b}$, we calculate, in the way similar to that in Section 3, the probability of both $\mathbf{a}$ and $\mathbf{b}$ satisfying a random instance $I \in \mathcal{L}$. Let m be the number of such indices i that $a_i = b_i$, there are also two cases:

$*$ either, $\mathbf{a}$ and $\mathbf{b}$ agree with each other on every variable of the constraint, in this case, the conditional probability of both $\mathbf{a}$ and $\mathbf{b}$ satisfying the constraint relation R_i is

$$\frac{d^k-1}{d-1} \Big/ \frac{(d^k-1)d}{d-1} = \frac{1}{d} = p;$$

$*$ or, the conditional probability of $\mathbf{a}$ and $\mathbf{b}$ satisfying constraint relation R_i is

$$\frac{d^{k-1}-1}{d-1} \Big/ \frac{(d^k-1)d}{d-1} = p\left(\frac{pd^k-1}{d^k-1}\right),$$

where $p = \frac{1}{d}$.

The probability that the first case occurs is $\sigma_{m,n} = \frac{\binom{m}{k}}{\binom{n}{k}}$. Thus we obtain that

$$\Pr(\mathbf{a} \in R_i, \mathbf{b} \in R_i) = p \cdot \sigma_{m,n} + p\left(\frac{pd^k-1}{d^k-1}\right)(1-\sigma_{m,n}).$$

Consequently we get

$$\Pr(S_{\mathbf{b}} = 1 | S_{\mathbf{a}} = 1) = \left(\sigma_{m,n} + \frac{pd^k-1}{d^k-1}(1-\sigma_{m,n})\right)^t.$$

Thus, all the arguments from the formula (7) to the end of Section 3 are still valid for completing a proof of Theorem 5.1. □

6. Conclusions

Motivated by k-SAT with growing k and based on the model RB, in this paper we proposed a model of random CSP. This model, named k-CSP, allows to deal with random CSP which domain size is fixed, tightness of constraint relations is fixed and length of constraints scopes grows very slowly. We proved theoretically the existence of phase transition in the model k-CSP and quantify the exact location of it. Experiments validate the effectiveness of the model k-CSP and illustrate that the computational complexity grows exponentially with the number of variables. Note that the worst-cases happen only around the phase transition point. In summary, since the experiments can be designed and conducted within a reasonably small range of domain size and constraint scope length and there is no other restrictions on the constraint relations except the fixed tightness, the model k-CSP can easily and naturally generate asymptotically non-trivial CSP instances and very suitable for testing the capability of CSP algorithms.

Combining the advantages of the proposed model k-CSP and linear CSP, we introduced a new type of random linear CSP model, named k-hyper-$\mathbb{F}$-linear CSP, by incorporating linear structure to the domains and constraint relations of the model k-CSP. Similar to the arguments established for the model k-CSP, the existence and exact location of phase transition are also demonstrated for the linear CSP model.

The investigation of random CSP in this paper is by no means exhaustive. As in the model k-CSP, this is suggestive of a more general random CSP model. The core concept of general-model-building is that of a certain relation between the size of domains and the size of constraints, including the length of constraint scopes and the tightness of constraint relations. Given such a relation, it is very possible to construct a more generalized model that has phase transition existing and is feasible to quantify it theoretically and more application extensively. This motivates our next step work to correlate the domain size with the constraint size to construct a relatively general CSP model.

Acknowledgements

The authors are grateful to Professor Ke Xu from Beihang University for his encouragement and many helpful discussions. Many thanks are given to the anonymous reviewers for their valuable comments which have helped us to improve this paper greatly.

This work is supported by National Science Project (973 Project) of China Grant No. 2005CB321902 and National Natural Science Foundation of China Grant No. 60973033.

Appendix A

Take A to be a finite index set, and let X_a for $a \in A$ be 0–1 random variables; then $X = \sum_{a\in A} X_a$ is a non-negative integer random variable. In Section 3 for the case that $r < 1$ we cite the following inequality

$$\Pr(X > 0) \geqslant \sum_{a\in A} \frac{\Pr(X_a = 1)}{\mathrm{E}(X|X_a = 1)},$$

which can be derived from the Jensen's inequality, see [25, Theorem 6.10]. For the model k-CSP, our proof in Section 3 by using the above inequality is essentially equivalent to the so-called second moment method, this can be seen from the formulas (2), (8) in Section 3 and the following lemma.

Lemma A.1. *Let $X = \sum_{a\in A} X_a$ be as above. If $\mathrm{E}(X_a) = \mathrm{E}(X_b)$ for any $a, b \in A$, then*

$$\sum_{a\in A} \frac{\Pr(X_a = 1)}{\mathrm{E}(X|X_a = 1)} \geqslant \frac{\mathrm{E}(X)^2}{\mathrm{E}(X^2)};$$

the equality holds if and only if $\mathrm{E}(X|X_a = 1) = \mathrm{E}(X|X_b = 1)$ for any $a, b \in A$.

Proof. Set $E = \mathrm{E}(X_a)$ for $a \in A$. Note that $X_b X_a$ is still a 0–1 random variable, hence $\Pr(X_b = 1, X_a = 1) = \Pr(X_b X_a = 1) = \mathrm{E}(X_b X_a)$. We have

$$\begin{aligned} \mathrm{E}(X|X_a = 1) &= \sum_{b\in A} \mathrm{E}(X_b|X_a = 1) = \sum_{b\in A} \Pr(X_b|X_a = 1) \\ &= \sum_{b\in A} \frac{\Pr(X_b = 1, X_a = 1)}{\Pr(X_a = 1)} = \sum_{b\in A} \frac{\mathrm{E}(X_b X_a)}{E}; \end{aligned}$$

so

$$\frac{\Pr(X_a = 1)}{\mathrm{E}(X|X_a = 1)} = \frac{E}{\sum_{b\in A} \frac{\mathrm{E}(X_b X_a)}{E}} = \left(\sum_{b\in A} \frac{\mathrm{E}(X_b X_a)}{E^2}\right)^{-1}.$$

By the inequality of arithmetic and harmonic means, we have

$$\begin{aligned}\sum_{a\in A}\frac{\Pr(X_a=1)}{\mathrm{E}(X|X_a=1)} &= \sum_{a\in A}\left(\sum_{b\in A}\frac{\mathrm{E}(X_bX_a)}{E^2}\right)^{-1}\\ &\geqslant |A|^2\cdot\left(\sum_{a\in A}\sum_{b\in A}\frac{\mathrm{E}(X_bX_a)}{E^2}\right)^{-1}\\ &= \frac{|A|^2E^2}{\sum_{a,b\in A}\mathrm{E}(X_bX_a)} = \frac{\sum_{a,b\in A}\mathrm{E}(X_b)\mathrm{E}(X_a)}{\sum_{a,b\in A}\mathrm{E}(X_bX_a)}\\ &= \frac{(\sum_{a\in A}\mathrm{E}(X_a))^2}{\mathrm{E}(\sum_{a,b\in A}X_aX_b)} = \frac{\mathrm{E}(X)^2}{\mathrm{E}(X^2)}.\end{aligned}$$

The equality holds if and only if for any $a, b \in A$ we have $\frac{\Pr(X_a=1)}{\mathrm{E}(X|X_a=1)} = \frac{\Pr(X_b=1)}{\mathrm{E}(X|X_b=1)}$, i.e. $\mathrm{E}(X|X_a = 1) = \mathrm{E}(X|X_b = 1)$. □

References

[1] D. Achlioptas, L.M. Kirousis, E. Kranakis, D. Krizanc, M.S.O. Molloy, Y.C. Stamatiou, Random constraint satisfaction: a more accurate picture, in: Proc. of the Third International Conference on Principles and Practice of Constraint Programming, in: LNCS, vol. 1330, 1997, pp. 107–120.

[2] D. Achlioptas, A. Naor, Y. Peres, Rigorous location of phase transitions in hard optimization problems, Nature 435 (7043) (2005) 759–764.

[3] G. Balakin, V. Kolchin, V. Khokhlov, Hypercycles in a random hypergraph, Diskretnaya Matematika 3 (3) (1991) 102–108.

[4] A. Bulatov, P. Jeavons, A. Krokhin, Classifying the complexity of constraints using finite algebras, SIAM Journal on Computing 34 (3) (2005) 720–742.

[5] N. Calkin, Dependent sets of constant weight vectors in GF(q), Random Structures and Algorithms 9 (1–2) (1996) 49–53.

[6] N. Calkin, Dependent sets of constant weight binary vectors, Combinatorics, Probability and Computing 6 (3) (1997) 263–271.

[7] P. Cheeseman, B. Kanefsky, W. Taylor, Where the really hard problems are, in: Proceedings of IJCAI-91, 1991, pp. 331–337.

[8] N. Creignou, H. Daudé, Approximating the satisfiability threshold for random k-XOR-CNF formulas, Combinatorics, Probability and Computing 12 (2) (2001) 113–126.

[9] N. Creignou, H. Daudé, Satisfiability threshold for random XOR-CNF formulas, Discrete Applied Mathematics 96–97 (1999) 41–53.

[10] N. Creignou, H. Daudé, Random generalized satisfiability problems, in: Proceedings of SAT, 2002.

[11] N. Creignou, H. Daudé, Generalized satisfiability problems: minimal elements and phase transitions, Theoretical Computer Science 302 (1–3) (2003) 411–430.

[12] N. Creignou, H. Daudé, Combinatorial sharpness criterion and phase transition classification for random CSPs, Information and Computation 190 (2) (2004) 220–238.

[13] R. Darling, J. Norris, Structure of large random hypergraphs, Annals of Applied Probability (2005) 125–152.

[14] O. Dubois, Y. Boufkhad, J. Mandler, Typical random 3-sat formulae and the satisfiability threshold, in: Proc. of SODA'00, 2000, pp. 126–127.

[15] O. Dubois, J. Mandler, The 3-XOR-SAT threshold, in: Proceedings of the 43rd Annual IEEE Symposium on Foundations of Computer Science, 2002, pp. 769–778.

[16] A. Flaxman, A sharp threshold for a random constraint satisfaction problem, Discrete Mathematics 285 (1–3) (2004) 301–305.

[17] E. Friedgut, Sharp thresholds of graph properties, and the k-SAT problem, Journal of the American Mathematical Society 12 (1999) 1017–1054, with an appendix by Jean Bourgain.

[18] E. Friedgut, Hunting for sharp thresholds, Random Structures and Algorithms 26 (1–2) (2005) 37–51.

[19] A. Frieze, M. Molloy, The satisfiability threshold for randomly generated binary constraint satisfaction problems, Random Structures and Algorithms 28 (3) (2006) 323–339.

[20] A.M. Frieze, N.C. Wormald, Random k-SAT: A tight threshold for moderately growing k, in: Proceedings of the 5th International Symposium on Theory and Applications of Satisfiability Testing, 2002, pp. 1–6.

[21] Y. Gao, J. Culberson, Consistency and random constraint satisfaction models with a high constraint tightness, in: CP04, 2004, pp. 17–31.

[22] Y. Gao, J. Culberson, Consistency and random constraint satisfaction problems, Journal of Artificial Intelligence Research 28 (2007) 517–557.

[23] I.P. Gent, E. MacIntyre, P. Prosser, B.M. Smith, T. Walsh, Random constraint satisfaction: flaws and structure, Constraints 6 (4) (2001) 345–372.

[24] A.C. Kaporis, L.M. Kirousis, E.G. Lalas, The probabilistic analysis of a greedy satisfiability algorithm, in: Proc. of the 10th Ann Eur. Symp. on Algor., 2002, pp. 574–585.

[25] M. Mitzenmacher, E. Upfal, Probability and Computing: Randomized Algorithm and Probabilistic Analysis, Cambridge Univ. Press, Cambridge, 2005.

[26] M. Molloy, Models and thresholds for random constraint satisfaction problems, in: Proceedings of the Thirty-Fourth Annual ACM Symposium on Theory of Computing, 2002, pp. 209–217.

[27] P. Prosser, An empirical study of phase transitions in binary constraint satisfaction problems, Artificial Intelligence 81 (1996) 81–109.

[28] B.M. Smith, M.E. Dyer, Locating the phase transition in binary constraint satisfaction problems, Artificial Intelligence 81 (1996) 155–181.

[29] B.M. Smith, Constructing an asymptotic phase transition in random binary constraint satisfaction problems, Theoretical Computer Science 265 (1–2) (2001) 265–283.

[30] K. Xu, F. Boussemart, F. Hemery, C. Lecoutre, Random constraint satisfaction: Easy generation of hard satisfiable instances, Artificial Intelligence 171 (2007) 514–534.

[31] K. Xu, W. Li, Exact phase transitions in random constraint satisfaction problems, Journal of Artificial Intelligence Reseach 12 (2000) 93–103.

[32] K. Xu, W. Li, Many hard examples in exact phase transition, Theoretical Computer Science 355 (2006) 291–302.