Open Problems in Theoretical Computer Science
China Theory Week 2012

Abstract. This document contains a list of open problems and related material that were posed during open problems sessions at China Theory Week 2012, held at Aarhus University in Aarhus, Denmark. Please send any comments/corrections regarding this document to dominik.scheder@gmail.com or joshua.e.brody@gmail.com.

Keynote Speakers.
Eric Allender
Christos Papadimitriou
Vijay Vazirani

Workshop Speakers.
Pablo Azar
Chris Beck
Jan Bulánek
Karl Bringmann
Gil Cohen
Radu Curticapean
Klim Efremenko
Jugal Garg
Sanjam Garg
Rasmus Ibsen-Jensen
Kasper Green Larsen
Kevin Lewi
Bruno Loff
Adriana Lopez
Ruta Mehta
Richard Peng
Hao Song
Alistair Stewart
Li-Yang Tan
Chengu Wang
Chris Wilkens
Yuli Ye

Question 1: Allocation of Divisible Goods Under Continuous Preferences (Vijay Vazirani)

There are N people and N beers. The goal is to allocate one liter of beer to each person. Of course, different people prefer different beers. We'd like a system that is incentive compatible, pareto optimal, and envy free.

One solution: each person has a pipe that drains one liter of beer per hour. Players place their pipe(s) into their favorite beer. When this beer is drained, they move to their next favorite beer.

Players have lexicographic preferences if their preferences are given by a permutation of the items, and a player (lexicographically)-prefers one allocation $a_i = (a_{i1}, \ldots, a_{im})$ to another allocation b_i if the most preferred item to the player that is not equally allocated in a_i and b_i is more in a_i.

Given lexicographic preferences, the greedy algorithm above satisfies several nice properties. It is incentive compatable, envy-free, efficient to compute, and pareto optimal.

Question: Is there a mechanism for allocation of divisible goods that has (approximately) the properties we'd like in a mechanism (incentive compatability, efficiency, envy-freeness, fairness) when players' preferences are represented by continuous utility functions?

Question 2: Finding a point outside a given subspace (Gil Cohen)

Input: a subspace U of $\mathbb{F}_2^n$ of dimension $n/2$.

Goal: Output in polynomial time a point x δ-far from U in Hamming distance.

Conjecture: $\delta = \Omega(n)$ should be possible.

Known: $\delta = \Omega(\log n)$ is possible.

The known solution is by Alon and others and works by covering the $O(\log n)$ neighborhood of U with small subspace objects. Then, find a point outside this space.

Question 3: Locally Decodable Codes (Klim Efremenko)

Known: There are three query LDCs linear over finite fields that have subexponential space.

Question (1): Are there three query LDCs over $\mathbb{C}$ or $\mathbb{R}$ of subexponential length?

Self-correctable codes: These are codes where you can recover any symbol of the code using a small number of queries.

Question (2): Does there exist self-correctable codes over $\mathbb{C}$?

Question 4: Cryptography by NP-Complete Problems (Sanjam Garg)

Consider a setting in which Alice has a public key and secret key pair (pk, sk). It publishes its public key pk and keeps its secret key sk private. Consider your favorite NP-complete problem (say Graph Hamiltonicity). Let G be an instance of the Graph Hamiltonicity problem. Bob has a message m and wants an encryption function $encrypt(m, pk, G)$ such that Alice with the knowledge of the Hamiltonian cycle in G (and her secret key sk) can decrypt in polynomial time. The security property desired is that no malicious probabilistic polynomial time (PPT) Alice should be able to figure out m if G is not Hamiltonian.

Question 5: Counting Contingency Matrices (Bruno Loff)

Let $\vec{\alpha} = (\alpha_1, \ldots, \alpha_d)$ and $\vec{\beta} = (\beta_1, \ldots, \beta_d)$ be integer vectors. A contingency matrix is a $d \times d$ nonnegative integer matrix such that the ith row sums to α_i and the jth column sums to β_j.

Counting the number of contingency matrices for given $\vec{\alpha}, \vec{\beta}$ is #-P complete.

There are upper bounds possible in $N^{d \log d}$ time.

Question (1) Are there any solutions over this time?

Question (2) Same problem, but with tensors. This is open even for $d = 3, 4$. For $d = 3$, best bound is $N^{d^2 \log(d)}$.

Question 6: Testing Monotonicity (Joshua Brody)

Testing Monotonicity. You have black-box access to a boolean function $f : \{0,1\}^n \to \{0,1\}$. How many queries do you need to distinguish f being monotone from f being ϵ-far from monotone?

f is monotone if $f(x) \leq f(y)$ whenever $x_i \leq y_i$ for all $1 \leq i \leq n$. Being ϵ-far from monotone means that you need to change an ϵ-fraction of the entries in the truth table of f to get a monotone function.

Question 7: Sensitivity vs. Average Sensitivity (Li-Yang Tan)

Let $f : \{0,1\}^n \to \{0,1\}$ be a boolean function, and $x \in \{0,1\}^n$. The sensitivity of f at x is $S_f(x) \# \{i \in [n] \text{ such that } f(x) \neq f(x^{\oplus i})\}$.

$AS(f) := \mathbb{E}_x[S_f(x)]$.

$S(f) := \max_x[S_f(x)]$.

It is obvious that $AS(f) \leq S(f)$. We conjecture that for monotone f, $AS(f)$ is in fact *much smaller* than $S(f)$:

Conjecture. *Let f be monotone and non-constant. Then $AS(f) \in O(S(f)^{0.99})$.*

Question 8: Hardness-on-Average for Subset Sum (Bruno Loff)

Let $S \subseteq \mathbb{Z}[x_1, \ldots, x_n]$ be a finite set of polynomials of degree d, $|S| = s$. The set $\{\mathbf{x} \mid p(\mathbf{x}) = 0 \ \forall \ p \in S\}$ partitions $\mathbb{R}^n$ into connected regions. There are roughly d^n such regions. Define the subset sum problem to be the following subset of $\mathbb{R}^n$:

$$\text{SSS} := \{\mathbf{x} \in \mathbb{R}^n \mid \exists \mathbf{a} \in \{0,1\}^n \text{ such that } \langle a, x \rangle = 1 \ .$$

One can show that SSS partitions $\mathbb{R}^n$ into 2^{n^2} many connected regions, even if we limit the bit length to be at most n. This means that SSS cannot be characterized by degree-d-polynomials.

Open Problem: Show hardness on average, that is, given any classification by polynomials of low degree, show that a positive fraction of SSS is wrongly classified (under an appropriate discretization of measure).

Question 9: Local Optima for Max-Cut (Rasmus Ibsen-Jensen)

Let $G = (V, E)$ be a bipartite graph with integer, but possibly negative, edge weights. It is hard to find a maximum cut of G. We want to find a *local optimum*, meaning that we cannot improve the value of the cut $V = U_1 \cup \bar{U}_1$ by moving a single vertex from U_1 to $\bar{U}_1$ or vice versa.

This problem can be seen to be equivalent to the following problem about matrices: Let $A \in \mathbb{Z}^{n \times n}$. We can now multiply entire rows and entire columns by -1, and our goal is to obtain a matrix in which all row sums and all column sums are non-negative. More formally, for $S \subseteq [n]$,

let I_S be the matrix whose entries are 0 off the diagonal, and the entry at (i,i) is -1 if $i \in S$ and 1 otherwise. Then the problem reads as follows:

Open Problem: Can we find, in polynomial time, two sets $S, T \subseteq [n]$ such that $I_S A I_T$ has non-negative row sums and column sums?

Note that the greedy algorithm, which multiplies a row or column by -1 if its sum is negative, terminates (the total sum of entries strictly increases in every step), thereby proving that such sets S, T always exist. However, its running time may be exponential.

Also, if we want local optimality under a neighborhood of radius 2, i.e., moving up to *two* vertices cannot improve the value, the problem becomes NP-hard.

Question 10: Strategy-Proof Facility Location (Dominik Scheder)

The problem is about facility location and mechanism design. There are n players, each living at a point p_i in some metric space X. The players report their locations to a central authority, which builds $k \leq n$ facilities $F = \{f_1, \ldots, f_k\}$ to serve the players. The cost of a player is

$$\text{cost}(i, F) := \min\{\text{dist}(p_i, f) \mid f \in F\} \ ,$$

and the social cost is

$$\text{cost}(F) := \sum_{i=1}^{n} \text{cost}(i, F) \ .$$

However, the players might report locations $q_1, \ldots, q_n$ that are different from their true locations, in order to improve their individual costs.

We want a mechanism that is *strategy proof*. That is, truthfully reporting $q_i = p_i$ is a dominant strategy for each player. Further, the mechanism should achieve a good approximation ratio, i.e., it should return an allocation F such that $\text{cost}(F)$ is not much larger than the cost of an optimal allocation F^*. "Not much larger" here can mean several things: From most to least satisfactory, it can mean "at most C times as much", "at most $C(k)$ times as much", or least "$C(k,n)$ times as much". It is known that already for $k = 2$, no deterministic mechanism achieves a constant approximation ratio, i.e., independent of n. There is, however, a simple *randomized* mechanism: Choose f_1 uniformly at random from $\{q_1, \ldots, q_n\}$, and then choose f_2 randomly, but with probabilisties proportional to $\text{dist}(q_i, f_1)$, i.e., players who are further away are more likely to be chosen. For $k \geq 3$, the natural generalization of this mechanism is not strategy proof anymore.

Open Problem: Give a randomized mechanism for $k \geq 3$ that is strategy proof and achieves any bounded approximation ratio, i.e., even $C(k,n)$ would be interesting.

Question 11: Degree of polynomials $p : [n] \to [m]$ (Gil Cohen)

We want a non-constant polynomial $p : [n] \to [m]$. How small can we make its degree? For $n = m$, the identity $p(x) = x$ shows that degree 1 works. For $m = n - 1$ we actually need degree $n - o(n)$. For $m > n$ the problem looks trivial, since the identity is of course also a function $p : [n] \to [m]$. The degree of p exhibits a gap, though:

Theorem 1. *If $p : [n] \to [n^d]$, then either $\deg p \leq d$ or $\deg p \geq \frac{n}{3} - \log n$.*

For an upper bound, there are polynomials $p : [n] \to [n]^d$ of degree $n - \log n$.

Conjecture: The gap in the theorem is actually between d and $n - o(n)$.

Question 12: Polynomial Restrictions with Low Variance (Li-Yang Tan)

Let $p := \{-1, 1\}^n \to [-1, 1]$. We can write is as a polynomial, also called the Fourier transformation of p:

$$p(x) = \sum_{S \subseteq [n]} \hat{p}_S \prod_{i \in S} x_i \ .$$

The coefficients $\hat{p}_S \in \mathbb{R}$ are called the Fourier coefficients of p. Let d be the degree of p, i.e., $\max |S|$ over all $S \subseteq [n]$ for which $\hat{p}_S \neq 0$.

Conjecture: There exists a restriction ρ of poly(d) many variables such that the variance of p_ρ is at most $1/10$. Here, the variance is taken with respect to the uniform distribution over $\{-1, 1\}^n$.

Question 13: Concentration of Influence (Chris Beck)

Suppose $T_1, \ldots, T_m$ are decision trees over n variables. Let X be the random variable counting the number of trees that output 1, for $x \in \{0, 1\}^n$ uniformly at random.

Theorem 2 (Stated very informally)**.** *If (1) all decision trees are of bounded height and (2) each variable i is on expectation read by only a small number of trees, then X is exponentially concentrated around its mean.*

Open Problem: Generalize the above result from decision trees to arbitrary boolean functions. The condition "every variable is on expectation read by only a small number of tress" should now become "the influence of each variable, averaged over all m functions, is small".

Here, the influence of a variable i is the probability, over x, that $f(x) \neq f(x + e_i)$, i.e., that flipping the i^{th} bit changes the function value.

Question 14: ϵ-Approximate Nash and 3-Player Games (Kristoffer Arnsfelt Hansen)

Computing a Nash equilibrium in a two-player game is PPAD-complete. Deciding whether a give has a Nash equilibrium of social welfare at least θ is even NP-hard. However, computing an ϵ-Nash equilibrium can be done in quasipolynomial time, to be precise, in time $2^{\text{poly}(\log(n), 1/\epsilon)}$.

Consider the following 3-player games: The payoff is given by a 3-dimensional tensor $a_{i,j,k}$, and the players are Irene, Jorge, and Karl, who choose actions i, j, and k respectively. Irene and Jorge are allies and want to maximize $a_{i,j,k}$, whereas Karl wants to minimize. We are interested in

$$\max \max \min(A) := \max_{\sigma_I, \sigma_J} \min_{\sigma_K} \mathbb{E}[a_{i,j,k}] \tag{1}$$

where σ_I, σ_J and σ_K are the mixed strategies of Irene, Jorge, and Karl, respectively. Note that this is, in general, different from $\min \sigma_K \max_{\sigma_I, \sigma_J} \mathbb{E}[a_{i,j,k}]$. Computing the value of (1) is NP-hard, and in particular there is a reduction from deciding whether a 2-player game has a Nash equilibrium of

social welfare at least θ.

Open Problem: To which extend does this reduction carry over to approximate versions of the stated problems?

The following is known: There is a reduction from 2-player games G to 3-player games H such that (i) if G has an ϵ-approximate Nash equilibrium of social welface at least $2(1-\alpha)$, then the $\max\max\min(H) \leq \alpha$; (ii) if all 2ϵ-approximate Nash equilibria of G achieve a social welfare of at most $2(1-\alpha-\epsilon)$, then $\max\max\min(H) \geq \alpha +$
epsilon.

Open Problem: Can you improve this reduction? In particular, can we replace "2ϵ-approximate" in (ii) by "ϵ-approximate"?