Asymptotic Behaviour of the Quadratic Knapsack Problem

Joachim Schauer

Department of Statistics and Operations Research

University of Graz, Austria

AGTAC - Koper 17.06.2015

Introduction

Quadratic Knapsack Problem (QKP)

Standard Knapsack Problem (KP) with additional "profits" p_{ij} for every pair of selected items i and j.

$$(QKP)\ \max \sum_{i=1}^{n}\sum_{j=1}^{n} p_{ij}x_ix_j \tag{1}$$

$$\text{s.t.} \sum_{i=1}^{n} w_ix_i \leq c \tag{2}$$

$$x_i \in \{0,1\}, \quad i = 1,\dots,n \tag{3}$$

$x_i = 1$ iff item i is included in the solution

surveys: Pisinger [2007], Kellerer et al. [2004] ch.12

Introduction

Graph Representation

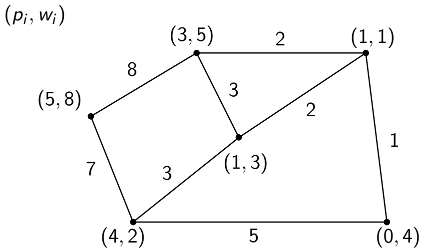
Usually, not all pairs (i, j) contribute quadratic profits.
Consider graph $G = (V, E)$ with $|V| = n$ and $|E| = m$.

- Every vertex $v \in V$ corresponds uniquely to an item.
- Edge $(u, v) \in E \iff$ two items corresponding to u, v yield an additional profit, if they are both included in the solution.

$$(QKP) \ \max \sum_{(i,j)\in E} p_{ij} x_i x_j \tag{4}$$

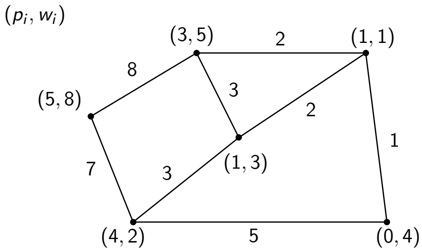
$x_{ii} \approx$ linear profit!

Example



$c = 11$

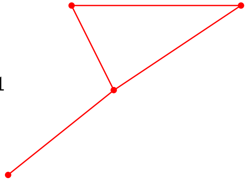
Example



$c = 11$

$P = 19$

$W = 11$



Applications and Solution Approaches

Applications

- media mix optimization (Pferschy and Sch. [2015])
- airport and train-station location (Rhys [1970])
- VLSI-design (Ferreira et al. [1996])
- ...

Exact Methods

- Caprara et al. [1999]: branch and bound based on Lagrangian relaxation
- Billionnet and Soutif [2004]: Lagrangian decomposition
- Pisinger et al. [2007]: aggressive reduction strategy in order to fix some variables
- Fomeni et al. [2014]: cut and branch for sparse instances

Solution Approaches

(Meta)-Heuristics

- Julstrom [2015,2012]: genetic algorithm
- Fomeni and Letchford [2014]: dynamic program combined with local search
- Yang et al. [2013]: tabu search and Grasp

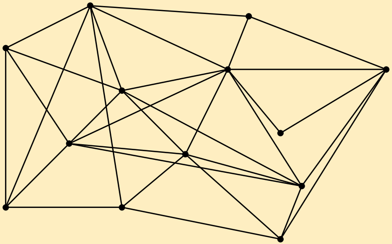
All these methods perform very well!

Yang et al. [2013] solve instances of up to 2000 items (gap $< 1.5\%$).

Known Hardness

- QKP is $\mathcal{NP}$ hard because of an easy reduction from maximum clique
- No hardness results under "standard" assumptions

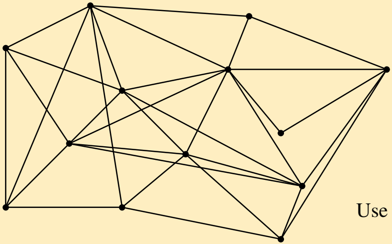
This result does not contradict the good results from above.



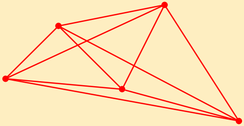
Known Hardness

- QKP is $\mathcal{NP}$ hard because of an easy reduction from maximum clique
- No hardness results under "standard" assumptions

This result does not contradict the good results from above.



$$p_{ii} = 0$$
$$w_i = 1$$
$$p_{ij} = 1 \ \forall (i,j) \in E$$



Use binary search for c!

Important Connections

Densest k-subgraph (dks)

GIVEN: graph $G = (V, E)$ and an integer k

FIND: k-vertex induced subgraph with most edges

Find the k vertex induced subgraph of a given graph $G = (V, E)$ containing the maximum number of edges.

It is obviously a subproblem of QKP.

Important Connections

Hardness results for dks

- Feige [2002] and Khot [2006] ruled out existence of a PTAS (average case hardness assumptions)
- Alon et al. [2011] ruled out any constant factor approximation (based on hardness of random k-AND formulas)
- Alon et al. [2011] showed superconstant inapproximation results (based on the hidden clique assumption)

Consequences for QKP

Hardness of QKP

All these results hold for QKP

Hence the empirically observed performance of the above algorithms raises questions:

- Are these (non-standard) complexity assumptions wrong?
- Is there something wrong with the algorithms, resp. with the instances used for testing them?

We will show that the used test-instances are problematic and give a new class of hard test-instances.

Test instances for QKP

Standard instances for QKP are randomly generated instances.
This is common for many optimization problems!

Instances by Gallo et al. [1980]

- a density Δ stands for the probability that a p_{ij} is non-zero
- whenever p_{ij} is non-zero, p_{ij} is uniformly distributed $\in [1, 100]$
- w_i is uniformly distributed $\in [0, 50]$
- c is uniformly distributed $\in [0, \sum w_i]$

These instances where used in all subsequent computational papers as core test instances.

Related Results for Quadratic Objectives

Quadratic assignment problem

$$\min_{\phi \in S_n} \left(\sum_{i=1}^{n} \sum_{k=1}^{n} a_{ik} b_{\phi(i)\phi(k)} + \sum_{i=1}^{n} c_{i\phi(i)} \right)$$

- n facilities are placed to n locations
- $c_{i\phi(i)}$ is the cost of opening facility i at location $\phi(i)$
- $a_{ik} b_{\phi(i)\phi(k)}$ is the transportation cost caused by assigning facility i to $\phi(i)$ and facility k to $\phi(k)$

Note that any feasible solution corresponds to a permutation of $\{1, 2, \ldots, n\}$.

Related Results for Quadratic Objectives

Asymptotic Result

Burkard and Frieze [1982] proved that:

- whenever the costs are $i.i.d$ random variables $\in [0, 1]$
- the ratio of the optimal and worst solution tends to 1 in probability

Generic Optimization Problems

Burkard and Frieze [1985] generalized this result to a broader class of optimization problems with quadratic objective.

They have in common that a feasible solution has a fixed number of n elements.

This does not hold for QKP - the empty knapsack is feasible.

Prerequisites

Chernoff-Hoeffding bound by Angluin and Valiant [1979]

Let the random variables $X_1, X_2, \ldots, X_n$ be independent with $0 \leq X_k \leq 1$ for each k. Let $S_n = \sum X_k$ and let $\mu = E(S_n)$. Then for any $0 \leq \varepsilon \leq 1$:

$$P\left[S_n \geq (1+\varepsilon)\mu\right] \leq e^{-\frac{1}{3}\varepsilon^2\mu}$$

$$P\left[S_n \leq (1-\varepsilon)\mu\right] \leq e^{-\frac{1}{2}\varepsilon^2\mu}$$

Prerequisites

Assumptions

- p_{ij} are $i.i.d.$ random variables defined on the interval $[0, 1]$
- weights are arbitrary numbers from $[0, 1]$
- the knapsack capacity c is proportional to n (i.e. $c = \lambda n$)
- all random variables have positive expectation (i.e. $E(X) = \mu_X > 0$).

asymptotic-QKP(n) problem:

a-*QKP*(*n*)

$$a\text{-}QKP(n) \qquad \max \sum_{i=1}^{n}\sum_{j=1}^{n} P_{ij}x_i x_j \tag{5}$$

$$\text{s.t.} \sum_{i=1}^{n} W_i x_i \leq \lambda n \tag{6}$$

$$x_i \in \{0,1\}, \quad i = 1, \ldots, n \tag{7}$$

If the weights are random variables:

L denotes the maximum number of items which can be feasibly included into the knapsack

L itself is a random variable

asymptotic-QKP(n) problem:

a-*QKP*(n)

Let a realization of W_i be given:

Then the realization of L can be determined by ordering the items in non-decreasing order of their realized weights.

$L \approx l$ such that $\sum_{i=1}^{l} w_i \leq \lambda n$ and $\sum_{i=1}^{l+1} w_i > \lambda n$.

Different Solutions

- $Z^A(n)$ denotes the random variable corresponding to the objective value that results by including the L lightest items.
- $Z^*(n)$ denotes the random variable which corresponds to the optimal solution value of the given instance.

Main Result

For any positive constant δ we get:

$$\lim_{n \to \infty} P\left[\frac{Z^*(n)}{Z^A(n)} \leq (1+\delta)\right] = 1$$

Hence the objective value of this easy heuristic converges in probability to the optimal objective value.

Consequences

- Testing QKP (meta)-heuristics with randomly generated instances is definitely not a good idea.
- Testing exact QKP algorithms with randomly generated instances should be done in a very careful way.

Sketch of Proof

Relax a-$QKP(n)$

Relax a a-$QKP(n)$ instance I by replacing the knapsack constraint with a cardinality constraint.

F_n^l denote set of all subsets of cardinality l ($|F_n^l| = \binom{n}{l} < 2^n$)

For a set S we define the objective value:

$$Z_I^S(n) = \sum_{i,j \in S} P_{ij}$$

Relaxed problem seeks for:

$$Z_I^{max}(n) = \max_{S \in F_n^l} Z_I^S(n) \qquad Z_I^{min}(n) = \min_{S \in F_n^l} Z_I^S(n)$$

Sketch of Proof

Crucial Observation

In an *a-QKP*(n) instance with n items at least λn items fit, hence $L \geq \lambda n$.

$Z^*(n)$ corresponds to a solution containing $\leq L$ items, hence there always exits a certain index $l' \geq \lambda n$ such that the following inequality holds:

$$Z_{l'}^{max}(n) \geq Z^*(n) \geq Z^A(n) \geq Z_{l'}^{min}(n) \qquad (8)$$

Sketch of Proof

By the linearity of expectation we get for all $S \in F_n^l$:

$$E[Z_l^S(n)] = E\left[\sum_{i \in S} P_{ii} + \sum_{1 \leq i < j \leq n | i,j \in S} P_{ij}\right] \geq \tag{9}$$

$$\geq l\mu_m + \frac{l(l-1)}{2}\mu_m \geq \frac{\lambda^2 n^2}{2}\mu_m \tag{10}$$

Sketch of Proof

Continuous Mapping

Show that for all $l \geq \lambda n$ the following holds:

$$\lim_{n \to \infty} P\left[\frac{Z_l^{max}(n)}{Z_l^{min}(n)} \leq (1+\delta)\right] = 1 \tag{11}$$

By the continuous mapping theorem it is enough to show that:

$$\lim_{n \to \infty} P\left[Z_l^{max}(n) \geq (1+\varepsilon)E(Q_l^n)\right] = 0 \tag{12}$$

$$\lim_{n \to \infty} P\left[Z_l^{min}(n) \leq (1-\varepsilon)E(Q_l^n)\right] = 0 \tag{13}$$

Sketch of Proof

Continuous Mapping

$E(Q_l^n)$ denotes the expected objective value over all knapsacks containing l items, while ignoring the capacity constraint:

$$E(Q_l^n) = \sum_{S \in F_n^l} \frac{Z_l^S(n)}{\binom{n}{l}} \tag{14}$$

Sketch of Proof

Equation (12): let S' now be any set of l knapsack items.

$$P\left[Z_l^{max}(n) \geq (1+\varepsilon)E(Q_l^n)\right] = P\left[\bigvee_{S \in F_n^l} \left(Z_l^S(n) \geq (1+\varepsilon)E(Q_l^n)\right)\right] \leq \tag{15}$$

$$\leq \sum_{S \in F_n^l} P\left[Z_l^S(n) \geq (1+\varepsilon)E(Q_l^n)\right] = |F_n^l| \cdot P\left[Z_l^{S'}(n) \geq (1+\varepsilon)E(Q_l^n)\right] \leq \tag{16}$$

$$\leq |F_n^l| \cdot e^{-\frac{1}{3}\varepsilon^2 E(Q_l^n)} \leq |F_n^l| \cdot e^{-\frac{1}{3}\varepsilon^2 \frac{\lambda^2 n^2}{2} \mu_m} \tag{17}$$

Remarks

Remarks

- The second inequality follows analogously.
- Almost sure convergence can be shown by applying the Borel-Cantelli-Lemma!
- The result not only covers the instances by Gallo et al. [1980], but a broad class of randomly generated instances.

Hidden Clique Problem

Erdos-Renyi Random Graph

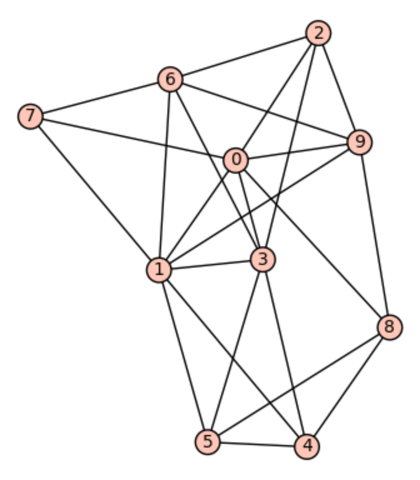
- $G(n, \frac{1}{2})$ has n vertices.
- each edge appears with probability $\frac{1}{2}$
- Fact: maximum clique has size $\approx 2\log_2(n)$

Hidden Clique Problem

- Plant a clique of size $l >> 2\log_2(n)$ into $G(n, \frac{1}{2})$
- Goal: find the planted clique

Hidden Clique Problem

Example of a $G(10, \frac{1}{2})$



Hidden Clique Assumption

Finding the hidden clique in polynomial time when $l = n^c$ with $c < \frac{1}{2}$ is impossible.

- Note that l is huge compared to the existing clique in $G(n, \frac{1}{2})$.
- The dks hardness result of Alon et al. [2011] is based on this assumption.

Hidden Clique Assumption

Hidden Clique instances

Plant a clique of size $\lfloor n^{\frac{1}{2}} \rfloor$ into a $G(n, \frac{1}{2})$:

- $p_{ii} = 0$ $w_i = 1$ and $p_{ij} = 1$ whenever $(i, j) \in E$
- $c = \lfloor n^{\frac{1}{2}} \rfloor$
- The optimal solution value of such an instance is (with overwhelming probability)

$$\frac{\lfloor n^{\frac{1}{2}} \rfloor \left(\lfloor n^{\frac{1}{2}} \rfloor - 1 \right)}{2}$$

Computational Results

For each size 10 randomly generated hidden clique instances have been tested with algorithms from the literature:

$n = 200,\ c = 14,\ opt = 91$

Fomeni and Letchford [2014]	Julstrom [2005] GA	own GA
78.9	85.4	88.3

$n = 800,\ c = 28,\ opt = 378$

Fomeni and Letchford [2014]	Julstrom [2005] GA	own GA
298.2	325.1	356.1

Thank you for your attention!

Thank you for your attention!