P. J. ŠAFÁRIK UNIVERSITY
FACULTY OF SCIENCE
INSTITUTE OF MATHEMATICS
Jesenná 5, 040 01 Košice, Slovakia

K. Cechlárová and E. Jelínková

An efficient implementation of the equilibrium algorithm for housing markets with duplicate houses

IM Preprint, series A, No. 2/2010
February 2010

An efficient implementation of the equilibrium algorithm for housing markets with duplicate houses*

Katarína Cechlárová[1] and Eva Jelínková[2]

[1]Institute of Mathematics, Faculty of Science, P.J. Šafárik University, Jesenná 5, 040 01 Košice, Slovakia

email: katarina.cechlarova@upjs.sk

[2]Department of Applied Mathematics, Faculty of Mathematics and Physics, Charles University, Malostranské náměstí 25, 118 00 Praha, Czech Republic

email: eva@kam.mff.cuni.cz

Abstract. We propose an asymptotically optimal implemetation of the equilibrium algorithm for housing markets with duplicate houses. It is based on Tarjan's depth-first search algorithm for strongly connected components of a digraph.

Keywords: Housing market, Economic equilibrium, Algorithm
AMS classification: 91A12, 91A06, 68Q25

1 Introduction

A housing market consists of a finite set of agents and a finite set of houses. Each agent owns one house, considers some other houses acceptable and orders them according to their desirability. The aim of the game for each agent is to get the house he finds the best possible. This model was introduced by Shapley and Scarf in [5], where also the notion of the economic equilibrium in such markets was defined and its existence in all housing markets proved. An algorithm for finding an economic equilibrium, called the Top Trading Cycles (TTC for short) algorithm was attributed to Gale (see [5]). An asymptotically optimal implementation of the TTC algorithm was proposed in [1].

However, the above results hold on the assumption that each agent's house is unique. If some houses may be equivalent, Fekete, Skutella and Woeginger [3]

*The research has been partially supported by project 1M0021620838 of the Czech Ministry of Education (Jelínková) and VEGA grants 1/0035/09 and 1/0325/10 (Cechlárová). Part of work was done while Jelínková was visiting ICE-TCS, Reykjavík University, Iceland.

proved that it is NP-complete to decide the existence of an economic equilibrium. Equivalent houses must have equal price and they are naturally tied in the preference orders of all agents. The market constructed in the NP-completeness reduction in [3] uses some additional ties. If ties are not present, a polynomial algorithm for this problem was proposed by Cechlárová and Fleiner in [2]. Their algorithm represents the housing market by a directed graph whose vertices correspond to house types and arcs to agents. The algorithm performs a repeated search for strongly connected components of certain subgraphs, so without a further improvement its complexity will be $O(NM)$, where N is the number of agents and M the number of house types. We show that using an extension of Tarjan's depth-first search algorithm for strongly connected components of a digraph an $O(L)$ implementation can be obtained, with L denoting the total length of preference lists of all agents. This is asymptotically optimal, since the size of the input is $\Theta(L)$.

2 Description of the model

Let A be a set of N *agents*, H a set of M *house types*; houses of the same type are said to be *equivalent*. The *endowment function* $\omega : A \to H$ assigns to each agent the type of house he originally owns. (Notice that in the classical Shapley–Scarf housing market [5], $M = N$ and ω is a bijection.) Each agent a finds some house types *acceptable* and expresses their desirability in the form of preferences, i.e. a linear ordering $P(a)$ of the set $H(a)$ of acceptable house types. We assume that $\omega(a) \in H(a)$ and this is the least preferred house in $H(a)$ for each $a \in A$. For each agent $a \in A$ and each set of house types $F \subset H$ denote by $f_F(a)$ the unique most preferred (sometimes we say the *first-choice*) house type of the set F.

The N-tuple of preferences $(P(a), a \in A)$ will be denoted by $\mathcal{P}$ and called the *preference profile*.

The *housing market* is a quadruple $\mathcal{M} = (A, H, \omega, \mathcal{P})$.

If $S \subseteq A$, we denote by $\omega(S) = \{\omega(a); a \in S\} \subseteq H$ the set of house types owned by agents in S and $\mathcal{M}' = \mathcal{M} \backslash S = (A \backslash S, \omega(A \backslash S), \omega', \mathcal{P}')$ is a *submarket* of $\mathcal{M}$ if ω' and $\mathcal{P}'$ are restrictions of ω and $\mathcal{P}$ to $A \backslash S$ and $\omega(A \backslash S)$, respectively.

We say that a function $x : S \to H$ is an *allocation* on $S \subseteq A$ if there exists a bijection π on S such that $x(a) = \omega(\pi(a))$ for each $a \in S$. A pair (x, p), where x is an allocation on A and $p : H \to \mathbb{R}$ is a price function, is an *economic equilibrium* for market $\mathcal{M}$ if each agent $a \in A$ receives a house of the most preferred type from the *affordable* house types, i.e. $x(a) = f_F(a)$ where $F = \{h \in H : p(h) \leq p(\omega(a))\}$.

It is known that if (x, p) is an economic equilibrium for market $\mathcal{M}$ then $p(x(a)) = p(\omega(a))$ for each $a \in A$ (see e.g. [3, 2]).

The algorithm for the economic equilibrium in [2] is based on a representation of a housing market $\mathcal{M}$ as a digraph $G_{\mathcal{M}} = (V, E)$, called the *digraph of the* *housing market* $\mathcal{M}$, with vertices representing house types, i.e. $V = H$, and arcs representing agents in the following way: if $\omega(a) = h_i$ and $h_j = f_H(a)$ for an agent a, then there is an arc $e(a) = (h_i, h_j) \in E$.

In this paper, the following graph-theoretical terminology is used. A *digraph* may contain parallel arcs as well as loops. If $V' \subseteq V$, a subdigraph of G *induced* by V' is a digraph $G(V') = (V', E')$ where $E' = \{(i,j) \in E; i,j \in V'\}$. A *walk* in G is a sequence of vertices $(i_0, i_1, \ldots, i_k)$ such that $(i_j, i_{j+1}) \in E$ for each $j = 0, 1, \ldots, k-1$. We say that vertex j is a *successor* of a vertex i in G, if G contains a walk from i to j. A vertex j is a *son* of a vertex i in G, if $(i,j) \in E$. A vertex $i \in V$ is a *sink* in G, if it has no successors in G. A digraph G with vertex set V is *strongly connected*, if for each pair i,j of vertices of V there is a walk from i to j as well as a walk from j to i in G. A *strongly connected component* (SCC for short) of a digraph G is a maximal strongly connected subdigraph of G. The *condensation* $G^* = (V^*, E^*)$ of a directed graph G is obtained by merging the vertices of each SCC of G. For $i \in V$ we shall denote by i^* the vertex of G^* corresponding to the SCC of G containing i. A SCC is *sink*, if its corresponding vertex in the condensation is a sink. There are several algorithms to construct a condensation of a digraph, linear in the number of its arcs; we shall take as our starting point Tarjan's algorithm based on the depth-first search of a digraph [6].

Cechlárová and Fleiner in [2] proved that if an economic equilibrium (x,p) exists in $\mathcal{M}$ and D is a sink SCC in $G_{\mathcal{M}}$, then

- the prices of all house types corresponding to vertices of D are equal,
- all agents corresponding to arcs of D trade among themselves, so D is an Eulerian digraph, and
- if $G_{\mathcal{M}}$ contains an arc (u,v) with $u \notin V(D)$ and $v \in V(D)$, i.e. there is an agent a who owns a house h of a type different from all the house types in D but his first-choice house type belongs to D, then this house may not be affordable for a, in other words, price of u must be strictly smaller than the price of house types in D.

These observations lead to the following theorem [2]:

Theorem 1 *Let $\mathcal{M}$ be a housing market and let D be a sink SCC in $G_{\mathcal{M}}$. Then $\mathcal{M}$ admits an economic equilibrium if and only if D is Eulerian and the reduced market $\mathcal{M}' = \mathcal{M} \backslash E(D)$ also admits an economic equilibrium.*

This theorem applies to Shapley-Scarf markets without ties too. If there are no duplicate houses in $\mathcal{M}$, then each vertex of $G_{\mathcal{M}}$ corresponds to a unique house owned by just one agent, the out-degree of each vertex is 1 and the top strongly connected components are simply cycles (so they are trivally Eulerian) and an economic equilibrium always exists.

Theorem 1 leads to the following algorithm for finding an economic equilibrium (or showing that there is none) for a given housing market $\mathcal{M}$: First, create the house-type digraph $G_{\mathcal{M}}$ and take any sink SCC D in $G_{\mathcal{M}}$. If D is not Eulerian, then $\mathcal{M}$ does not admit any economic equilibrium. If D is Eulerian, the agents and house types corresponding to arcs and vertices of D are deleted from $\mathcal{M}$ and the whole procedure continues for the obtained submarket $\mathcal{M}' = \mathcal{M} \backslash V(D)$. However, $G_{\mathcal{M}'}$ is not simply a subgraph of $G_{\mathcal{M}}$, as in the digraph corresponding to $\mathcal{M}'$ some new arcs are added. They correspond to agents whose endowments are in $\mathcal{M}'$ but their first-choice house types are in $V(D)$.

A sink SCC in $G_{\mathcal{M}}$ can be found in $O(N)$ steps (since the arcs in $G_{\mathcal{M}}$ correspond to agents – better to say, to their unique first-choice house types). After its deletion, the market has lost at least one house-type, so the number of iterations is bounded by $O(M)$. In the digraph corresponding to the submarket, some arcs remain from the previous iteration, but some new may emerge (as said above). If we compute the condensation from scratch, we cannot ensure a better bound than $O(N)$ for the new iteration. So the complexity of the whole algorithm will be $O(NM)$.

The aim of this paper is to show how to extend Tarjan's depth-first search based algorithm for strongly connected components of a digraph [6] to accomplish this task in $O(L)$ time, where L is the total length of preference lists or agents of the market, i.e. $L = \sum_{a \in A} |H(a)|$.

3 Description of the algorithm

We shall start with Tarjan's algorithm as described in [4, 7]. This algorithm creates a rooted directed forest for the explored digraph G by visiting (procedure $Visit$) the vertices of G in the depth-first search (DFS for brevity) order. The vertices are labelled by their DFS numbers $numb$ according to this order. Each SCC is a subtree of the constructed forest, so to compute the condensation means to identify the roots of strongly connected components. To identify them, for each vertex v the number $low(v)$ is computed, which denotes the smallest DFS number of a vertex in the component containing v. A vertex v is a root of a SCC if and only if $numb(v) = low(v)$ (see Theorem 11.15 in [7]).

The visited vertices are stored in an auxiliary stack (for them the variable *stacked* is set to be *true*) so that the vertices of the SCC with root v were pushed onto the stack after v. As soon as such a SCC is discovered (this happens after all the sons of v have been visited), all its vertices together with v are popped from the stack and *stacked* is set to $false$ for them.

In the proof of the correctness of Tarjan's algorithm, the key role is played by proving the correctness of the computation of variables $low(v)$. To achieve this, the newly explored arcs (v, w) leading from vertex v are classified according to their types:

- If $numb(w) = 0$ (i.e. arc (v, w) leads to a currently unvisited vertex) then this arc is a *forest* arc. On taking such an arc, procedure $Visit$ is called for vertex w and vertex w is pushed onto the stack. When the search returns to v after all the vertices in the subtree rooted at w have been visited, $low(v)$ is set to minimum of its current value and $low(w)$ unless w was removed from stack as a root of a new SCC.

- If $numb(w) > numb(v)$ then (v, w) is a *forward* arc, leading to a DFS descendant of v. This arc has no influence on the outcome, since the DFS tree already contains a directed path from v to w.
- If $0 < numb(w) < numb(v)$ then (v, w) is either a *back* arc (leading to a DFS ancestor of v) or a *cross* arc (leading to a vertex that has already been visited, but is neither a descendant nor an ancestor of v). In this case $low(v)$ is modified to be the minimum of $low(v)$ and $numb(w)$, unless w is in a different SCC than v. The latter can only happen for cross arcs and this situation occurs if and only if $stacked(w) = false$.

The complete pseudocode of our modification of Tarjan's algorithm is given Figure 1. We started with the recursive definition of this algorithm in Figure 1.20 of [4] and incorporated two main changes:

- A cross arc or a forward arc e from v to w may lead to a SCC that was discovered before; this occurs if $stacked(w) = false$. (Notice that this cannot happen for a back arc, since then v and w are in the same SCC.) We shall call such arcs *invalid* (naturally, all the other arcs are *valid*). An invalid arc does not modify $low(v)$, but the agent that corresponds to it has to take his next-choice house type. Invalid arcs are discovered in line 8 (here the SCC containing vertex w has been encountered before the arc e is explored) and in line 15 (this happens immediately after w is identified as a root of a new SCC). In both cases, arc e is made invalid and it has to be replaced by the next-choice house type of the agent corresponding to e. This is indicated by statement Upgrade(e), which also means that the replacement arc is added into $Out(v)$.

  This dynamic nature of the sets of outgoing arcs for vertices also means that the simple **for** instruction is replaced by a **while** cycle (lines 5 up to 21) to work with the sons of a vertex.
- We need to check the Eulerian property for each discovered strongly connected component, correctly taking into account only arcs within this component.

The correctness of the described changes is justified in the following lemmas:

Lemma 1 *The only arcs that can change their status from valid in the course of algorithm HousingEquilibrium are the forest arcs.*

Proof. Indeed, a forest arc (v, w) is first treated as a valid arc (line 13); it may become invalid if after finishing the exploration of w the subtree rooted at w is deleted from the stack (line 15).

A back arc (v, w) can never be invalid, since v and w are in the same SCC.

Now take a forward or cross arc (v, w) that was first valid and suppose that it became invalid. This may have happened in the moment when the algorithm discovered a SCC u^* containing w, rooted at a vertex u. However, when the root of this SCC is discovered, the only arc leading to some vertex of u^* is the forest arc leading to u (all the other arcs leading into u^* will be discovered later and will already be invalid at the time of their discovery), so no forward of cross arc can change its status from valid to invalid. ∎

Lemma 2 *Algorithm HousingEquilibrium correctly identifies the sequence of strongly connected components according to Theorem 1.*

Proof. First, realize that until the first SCC is identified (notice that it must be a sink SCC), algorithm HousingEquilibrium works identically with Tarjan's algorithm. After some SCCs have been discovered, arcs leading to them are recognized and replaced by their next-choices correctly. Since each new arc introduced as an upgrade of an invalid arc can never start in an already deleted vertex neither lead to already deleted vertices, the argument is ready. ∎

Lemma 3 *Algoritm HousingEquilibrium correctly tests the Eulerian property of the obtained strongly connected components.*

Proof. We need to show that each arc (v, w) correctly contributes to $outdegree(v)$ and $indegree(w)$ in their respective SCCs. At the start of the algorithm, the values $outdegree(v)$ and $indegree(w)$ are set to 0 for each vertex. Invalid arcs do not modify these variables, but if the arc (v, w) is valid, we increase $outdegree(v)$ and $indegree(w)$ by 1. Since by Lemma 1 the only arcs that can change their status from valid to invalid are the forest arcs, we only have now to care about them. And indeed, if a forest arc (v, w) is deleted from the graph and upgraded (line 15), which happens when a new SCC with root w is discovered (line 22), $indegree(w)$ is decreased by 1. If this happens during searching the sons of vertex v, the $outdegree(v)$ is decreased by 1. Finally, to account for the SCCs whose root v is a root of a tree in the DFS forest (and hence there is no forest arc coming into v), we set $indegree(v)$ to 1 each time the search starts from a vertex v with $numb(v) = 0$. ∎

Theorem 2 *Algorithm HousingEquilibrium correctly decides the existence of an economic equilibrium in a housing market with duplicate houses in $O(L)$ steps, where L is the total length of all preference lists of agents.*

```
Input: a housing market M.
procedure Visit(v)
  begin
    numb(v) ← i; low(v) ← numb(v); i ← i + 1;
    push v on the stack; stacked(v) ← true;
    while Out(v) ≠ ∅ do
    begin
      take any e = (v, w) ∈ Out(v); delete e from Out(v);
      if numb(w) > 0 and not stacked(w)      (comment: invalid arc)
      then Upgrade(e);
      else
      begin
      outdegree(v) ← outdegree(v) + 1; indegree(w) ← indegree(w) + 1;
        if numb(w) = 0 then      (comment: forest arc)
          begin Visit(w);
            if not stacked(w) then      (comment: invalid arc)
            begin outdegree(v) ← outdegree(v) − 1; Upgrade(e) end
            else low(v) ← min{low(v); low(w)}
          end
        else if numb(w) < numb(v) then low(v) ← min{low(v); numb(w)}
      end;
    end;
    if  numb(v) = low(v) then      (comment: new SCC discovered)
    begin price ← price − 1;
      indegree(v) ← indegree(v) − 1;
      for all u in stack up to and including v do
      begin if indegree(u) ≠ outdegree(u) then equilibrium ← false and STOP;
        pop u from the stack; p(u) ← price; stacked(u) ← false
      end;
    end
  end
begin
  i ← 1; price ← m + 1; equilibrium ← true;
  empty the stack;
  for all v ∈ V do
  begin
    numb(v) ← 0; stacked(v) ← false; indegree(v) ← 0; outdegree(v) ← 0;
    put to Out(v) all first choices of agents of v
  end;
  while for some u ∈ V, numb(u) = 0 do begin indegree(v) := 1; Visit(u) end
end
```

Figure 1: Algorithm HousingEquilibrium

Proof. The correctness of HousingEquilibrium follows from Lemma 2, Lemma 3, and Theorem 1.

The digraph of the housing market has M vertices and $O(L)$ arcs (including those added by Upgrade). Since $M \leq L$, and since Tarjan's algorithm works in time linear in the size of the digraph, the time complexity bound $O(L)$ follows. ■

4 Conclusion

We have presented an asymptotically optimal algorithm HousingEquilibrium for computing the economic equilibrium in housing markets with duplicate houses.

Notice that this algorithm (with suitable simplifications, for example values of *outdegree* and *indegree* do not have to be computed) provides another asymptotically optimal implementation of the Top Trading Cycles algorithm as an alternative for the implementation described in [1].

References

[1] D. Abraham, K. Cechlárová, D. Manlove and K. Mehlhorn, *Pareto optimality in house allocation problems*, Lecture Notes in Comp. Sci. **3341**, Algorithms and Computation, ISAAC 2004, Hong Kong, December 2004, Eds. R. Fleischer, G. Trippen, 3-15 (2004).

[2] K. Cechlárová and T. Fleiner, *Housing markets through graphs*, Algorithmica, to appear, 2010.

[3] S. Fekete, M. Skutella, G. Woeginger, *The complexity of economic equilibria for house allocation markets*, Inform. Process. Lett. **88/5**: 219–223 (2003).

[4] A. M. Gibbons, *Algorithmic graph theory*, Cambridge University Press, Cambridge, 1985.

[5] L. Shapley and H. Scarf, *On cores and indivisibility*, J. Math. Econ. **1**:23–37 (1974).

[6] R. E. Tarjan, *Depth-first search and linear graph algorithms*, SIAM Journal on Computing **1(2)**:146–160 (1972).

[7] K. Thulasiraman and M. N. S. Swamy, *Graphs: theory and algorithms*, Wiley, 1992.

Recent IM Preprints, series A

2006

1/2006 Semanišinová I. and Trenkler M.: *Discovering the magic of magic squares*
2/2006 Jendroľ S.: *NOTE – Rainbowness of cubic polyhedral graphs*
3/2006 Horňák M. and Woźniak M.: *On arbitrarily vertex decomposable trees*
4/2006 Cechlárová K. and Lacko V.: *The kidney exchange problem: How hard is it to find a donor ?*
5/2006 Horňák M. and Kocková Z.: *On planar graphs arbitrarily decomposable into closed trails*
6/2006 Biró P. and Cechlárová K.: *Inapproximability of the kidney exchange problem*
7/2006 Rudašová J. and Soták R.: *Vertex-distinguishing proper edge colourings of some regular graphs*
8/2006 Fabrici I., Horňák M. and Jendroľ S., ed.: *Workshop Cycles and Colourings 2006*
9/2006 Borbeľová V. and Cechlárová K.: *Pareto optimality in the kidney exchange game*
10/2006 Harminc V. and Molnár P.: *Some experiences with the diversity in word problems*
11/2006 Horňák M. and Zlámalová J.: *Another step towards proving a conjecture by Plummer and Toft*
12/2006 Hančová M.: *Natural estimation of variances in a general finite discrete spectrum linear regression model*

2007

1/2007 Haluška J. and Hutník O.: *On product measures in complete bornological locally convex spaces*
2/2007 Cichacz S. and Horňák M.: *Decomposition of bipartite graphs into closed trails*
3/2007 Hajduková J.: *Condorcet winner configurations in the facility location problem*
4/2007 Kovárová I. and Mihalčová J.: *Vplyv riešenia jednej difúznej úlohy a následný rozbor na riešenie druhej difúznej úlohy o 12-tich kockách*
5/2007 Kovárová I. and Mihalčová J.: *Prieskum tvorivosti v žiackych riešeniach vágne formulovanej úlohy*
6/2007 Haluška J. and Hutník O.: *On Dobrakov net submeasures*
7/2007 Jendroľ S., Miškuf J., Soták R. and Škrabuláková E.: *Rainbow faces in edge colored plane graphs*
8/2007 Fabrici I., Horňák M. and Jendroľ S., ed.: *Workshop Cycles and Colourings 2007*
9/2007 Cechlárová K.: *On coalitional resource games with shared resources*

2008

1/2008 Miškuf J., Škrekovski R. and Tancer M.: *Backbone colorings of graphs with bounded degree*
2/2008 Miškuf J., Škrekovski R. and Tancer M.: *Backbone colorings and generalized Mycielski's graphs*
3/2008 Mojsej I.: *On the existence of nonoscillatory solutions of third order nonlinear differential equations*
4/2008 Cechlárová K. and Fleiner T.: *On the house allocation markets with duplicate houses*

5/2008 Hutník O.: *On Toeplitz-type operators related to wavelets*

6/2008 Cechlárová K.: *On the complexity of the Shapley-Scarf economy with several types of goods*

7/2008 Zlámalová J.: *A note on cyclic chromatic number*

8/2008 Fabrici I., Horňák M. and Jendroľ S., ed.: *Workshop Cycles and Colourings 2008*

9/2008 Czap J. and Jendroľ S.: *Colouring vertices of plane graphs under restrictions given by faces*

2009

1/2009 Zlámalová J.: *On cyclic chromatic number of plane graphs*

2/2009 Havet F., Jendroľ S., Soták R. and Škrabuľáková E.: *Facial non-repetitive edge-colouring of plane graphs*

3/2009 Czap J., Jendroľ S., Kardoš F. and Miškuf J.: *Looseness of plane graphs*

4/2009 Hutník O.: *On vector-valued Dobrakov submeasures*

5/2009 Haluška J. and Hutník O.: *On domination and bornological product measures*

6/2009 Kolková M. and Pócsová J.: *Metóda Monte Carlo na hodine matematiky*

7/2009 Borbeľová V. and Cechlárová K.: *Rotations in the stable b-matching problem*

8/2009 Mojsej I. and Tartaľová A.: *On bounded nonoscillatory solutions of third-order nonlinear differential equations*

9/2009 Jendroľ S. and Škrabuľáková E.: *Facial non-repetitive edge-colouring of semiregular polyhedra*

10/2009 Krajčiová J. and Pócsová J.: *Galtonova doska na hodine matematiky, kvalitatívne určenie veľkosti pravdepodobnosti udalostí*

11/2009 Fabrici I., Horňák M. and Jendroľ S., ed.: *Workshop Cycles and Colourings 2009*

12/2009 Hudák D. and Madaras T.: *On local propeties of* 1*-planar graphs with high minimum degree*

13/2009 Czap J., Jendroľ S. and Kardoš F.: *Facial parity edge colouring*

14/2009 Czap J., Jendroľ S. and Kardoš F.: *On the strong parity chromatic number*

2010

1/2010 Cechlárová K. and Pillárová E.: *A near equitable* 2*-person cake cutting algorithm*

Preprints can be found in: http://umv.science.upjs.sk/index.php/veda-a-vyskum/preprinty