

TOPICS  
IN  
DISCRETE  
MATHEMATICS

12

# COMPUTATIONAL MOLECULAR BIOLOGY

S. ISTRAIL  
P. PEVZNER  
R. SHAMIR  
(Editors)



NORTH-HOLLAND

# COMPUTATIONAL MOLECULAR BIOLOGY

*Edited by*

**Sorin Istrail**

*Celera Genomics, Rockville, MD, USA*

**Pavel Pevzner**

*University of California, San Diego, CA, USA*

**Ron Shamir**

*Tel Aviv University, Tel Aviv, Israel*

Reprinted from the journal

Discrete Applied Mathematics, Volume 127, Number 1 (2003)



2003

ELSEVIER

Amsterdam – Boston – London – New York – Oxford – Paris – San Diego  
San Francisco – Singapore – Sydney – Tokyo

ELSEVIER SCIENCE B.V.  
Sara Burgerhartstraat 25  
P.O. Box 211, 1000 AE Amsterdam, The Netherlands

© 2003 Elsevier Science B.V. All rights reserved.

This work is protected under copyright by Elsevier Science, and the following terms and conditions apply to its use:

#### Photocopying

Single photocopies of single chapters may be made for personal use as allowed by national copyright laws. Permission of the Publisher and payment of a fee is required for all other photocopying, including multiple or systematic copying, copying for advertising or promotional purposes, resale, and all forms of document delivery. Special rates are available for educational institutions that wish to make photocopies for non-profit educational classroom use.

Permissions may be sought directly from Elsevier's Science & Technology Rights Department in Oxford, UK: phone: (+44) 1865 843830, fax: (+44) 1865 853333, e-mail: [permissions@elsevier.com](mailto:permissions@elsevier.com). You may also complete your request online via the Elsevier Science homepage (<http://www.elsevier.com>), by selecting 'Customer Support' and then 'Obtaining Permissions'.

In the USA, users may clear permissions and make payments through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA; phone: (+1) (978) 7508400, fax: (+1) (978) 7504744, and in the UK through the Copyright Licensing Agency Rapid Clearance Service (CLARCS), 90 Tottenham Court Road, London W1P 0LP, UK; phone: (+44) 207 631 5555; fax: (+44) 207 631 5500. Other countries may have a local reprographic rights agency for payments.

#### Derivative Works

Tables of contents may be reproduced for internal circulation, but permission of Elsevier Science is required for external resale or distribution of such material.

Permission of the Publisher is required for all other derivative works, including compilations and translations.

#### Electronic Storage or Usage

Permission of the Publisher is required to store or use electronically any material contained in this work, including any chapter or part of a chapter.

Except as outlined above, no part of this work may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without prior written permission of the Publisher.

Address permissions requests to: Elsevier's Science & Technology Rights Department, at the phone, fax and e-mail addresses noted above.

#### Notice

No responsibility is assumed by the Publisher for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein. Because of rapid advances in the medical sciences, in particular, independent verification of diagnoses and drug dosages should be made.

Reprinted from: Discrete Applied Mathematics, Volume 127, Number 1 (2003)

#### Library of Congress Cataloging in Publication Data

A catalog record from the Library of Congress has been applied for.

#### British Library Cataloguing in Publication Data

A catalogue record from the British Library has been applied for.

ISBN: 0 444 51384 1

∞ The paper used in this publication meets the requirements of ANSI/NISO Z39.48-1992 (Permanence of Paper).  
Printed in The Netherlands.

# COMPUTATIONAL MOLECULAR BIOLOGY

Reprints from the journals *Discrete Mathematics* and  
*Discrete Applied Mathematics*

**General Editor: Peter L. Hammer**  
*Rutgers University, Piscataway, NJ, USA*

ELSEVIER

Amsterdam – Boston – London – New York – Oxford – Paris – San Diego  
San Francisco – Singapore – Sydney – Tokyo

## Contents

<i>S. Istrail, P. Pevzner and R. Shamir</i> Preface	1
<i>T. Akutsu, K. Kanaya, A. Ohyama and Asao Fujiyama</i> Point matching under non-uniform distortions	5
<i>C.L. Bajaj, V. Pascucci, A. Shamir, R.J. Holt and A.N. Netravali</i> Dynamic maintenance and visualization of molecular surfaces	23
<i>P. Damaschke</i> Point placement on the line by distance data	53
<i>F. Denis and O. Gascuel</i> On the consistency of the minimum evolution principle of phylogenetic inference	63
<i>I. Miklós</i> Algorithm for statistical alignment of two sequences derived from a Poisson sequence length distribution	79
<i>R. Pendavingh, P. Schuurman and G.J. Woeginger</i> Recognizing DNA graphs is difficult	85
<i>B. Schwikowski and M. Vingron</i> Weighted sequence graphs: boosting iterated dynamic programming using locally suboptimal solutions	95
<i>V. Veeramachaneni, P. Berman and W. Miller</i> Aligning two fragmented sequences	119
<i>V. Chandru, A. DattaSharma and V.S. Anil Kumar</i> The algorithmics of folding proteins on lattices	145
<i>V. Heun</i> Approximate protein folding in the HP side chain model on extended cubic lattices	163



ELSEVIER

Discrete Applied Mathematics 127 (2003) 1–3

---

---

DISCRETE  
APPLIED  
MATHEMATICS

---

---

## Preface

This is the fourth installment in the *Discrete Applied Mathematics* series on computational molecular biology, which is devoted to combinatorial and algorithmic techniques in computational molecular biology. This series publishes novel research results on the mathematical and algorithmic foundations of the inherently discrete aspects of computational biology. The previous volumes in the series were Volume 71, issue 1–3, December 1996, Volume 88, issues 1–3, November 1998, and Volume 104, issues 1–3, August 2000.

The current issue contains papers demonstrating the variety and richness of computational problems motivated by molecular biology. The application areas within biology that give rise to the problems studied in these papers include solid molecular modeling, sequence comparison, phylogeny, evolution, mapping, DNA chips, protein folding and 2D gel technology. The mathematical techniques used are algorithmics, combinatorics, optimization, probability, graph theory, complexity and applied mathematics. Below is a brief description of each paper.

In “Dynamic maintenance and visualization of molecular surfaces”, Bajaj, Pascucci, Shamir, Holt and Netravali study the problems of computing and updating the boundary representation of a molecular surface. Using non-uniform rational B-splines, they show how to efficiently compute the representation and how to gain speed in dynamic updates at the expense of accuracy.

Veeramachaneni, Berman and Miller, in their paper “Aligning two fragmented sequences”, study the problem of sequence comparison when each of the sequences is fragmented into several contigs whose order and orientation are unknown. This problem arises in analysis of “unfinished” DNA sequences. They show that optimal alignment is hard in this context, but provide a polynomial approximation algorithm.

In “Algorithm for statistical alignment of two sequences derived from a Poisson sequence length distribution”, Miklos provides an algorithm for computing the joint probability of two sequences evolved in a non-reversible way from a Poisson sequence length distribution.

Multiple sequence alignment is one of the fundamental challenges in computational biology. In “Weighted sequence graphs: Boosting iterated dynamic programming using locally suboptimal solutions”, Schwikowski and Vingron contribute to this field by developing a framework for iterated dynamic programming and applying to it benchmark alignment problems.

Sequencing by hybridization and universal DNA chips have ignited the imagination of theorists for over a decade, and have been a source of combinatorial and

algorithmic problems. In “Recognizing DNA graphs is difficult”, Pendavingh, Schuurman and Woeginger study the complexity of graphs motivated by SBH. In particular, they show that recognizing DNA graphs is NP-hard.

Denis and Gascuel, in “On the consistency of the minimum evolution principle of phylogenetic inference”, study a fundamental statistical question in phylogeny. To address the minimum evolution problem, they use a model that combines properties of ordinary and weighted least squares criteria and show that the minimum evolution principle is statistically consistent within this model.

In “Point matching under non-uniform distortions”, Akutsu, Kanaya, Ohyama and Fujiyama address a pattern matching problem that is related to two dimensional gel images analysis. They show that the problem is NP-hard in general, but provide an efficient solution for a special case and a practical heuristic.

In “Point placement on the line by distance data”, Damaschke studies a problem motivated by DNA mapping. The question is how to determine the placement of points on the line based on adaptive queries on their pairwise distances. The author proves upper and lower bounds on the number of necessary queries. He also provides an algorithm for generating all linear layouts given all pairwise distances of edges of a chordal graph.

Two papers deal with protein folding problems on lattices. In “The algorithmics of folding proteins on lattices”, Chandru, Datta Sharma and Anil Kumar survey the recent developments in the analysis of the problem, detailing both hardness and approximation results. In “Approximate protein folding in the HP side chain model on extended cubic lattices”, Heun studies the problem on a cubic lattice which is extended by diagonals in the plane. He provides two polynomial approximation algorithms that guarantee approximation ratios of  $59/70$  and  $37/42$ , respectively.

The vision of the “marriage” between mathematics and biology, which was also behind the idea of starting this series more than six years ago, is becoming a reality. In the “post-genome era” it is almost unthinkable that the biology of the future can be done without using computers for data handling and for sophisticated analysis. Biology continues to be a wonderful source of computational problems and challenges to the theorist and the practitioner in computer science and mathematics. In return, the theory developed on those problems leads to practical tools that advance the biological and medical research.

Sorin Istrail

*Celera Genomics 45 West Gude Drive  
Rockville, Md 20850  
USA*

*E-mail:* [sorin.istrail@celera.com](mailto:sorin.istrail@celera.com)

*URL:* <http://www.cs.sandia.gov/~scistra/>

Pavel Pevzner

*Department of Computer Science and Engineering  
University of California San Diego, APM 3132  
9500 Gilman Drive Dept. 0114*



*La Jolla, CA 92093-0114*

*USA*

*E-mail:* ppezvner@cs.ucsd.edu

*URL:* <http://www-cse.ucsd.edu/users/ppezvner/>

*Ron Shamir*

*School of Computer Science*

*Tel Aviv University*

*Tel Aviv 69978*

*Israel*

*E-mail:* rshamir@post.tau.ac.il

*URL:* <http://www.cs.tau.ac.il/~rshamir/>



# Point matching under non-uniform distortions<sup>☆</sup>

Tatsuya Akutsu<sup>a,\*</sup>, Kyotetsu Kanaya<sup>b</sup>, Akira Ohyama<sup>b</sup>,  
Asao Fujiyama<sup>c</sup>

<sup>a</sup>*Human Genome Center, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai,  
Minato-ku, Tokyo 108-8639, Japan*

<sup>b</sup>*Department of Bioscience Systems, Mitsui Knowledge Industry Co., Ltd. 2-7-14 Higashinakano,  
Nakano-ku, Tokyo 164-8555, Japan*

<sup>c</sup>*National Institute of Genetics, 1111 Yata, Mishima-city, Shizuoka 411-8540, Japan*

Received 28 October 2000; received in revised form 11 July 2001; accepted 4 September 2001

---

## Abstract

This paper discusses the pattern matching problem for points under non-uniform distortions, which arises from the analysis of two-dimensional (2-D) electrophoresis images. First, we provide a formal definition of the problem. Next, we prove that it is NP-hard in two (or more) dimensions. This proof is based on a reduction from planar 3SAT. Then we present a simple polynomial time algorithm for a special and one-dimensional case of the problem, which is based on dynamic programming. We also present a practical heuristic algorithm for identifying a match between two sets of spots in 2-D gel electrophoresis images obtained from genomic DNA. © 2002 Elsevier Science B.V. All rights reserved.

**Keywords:** Point matching; Geometric matching; Electrophoresis; NP-hard

---

## 1. Introduction

Matching of spatial point sets (i.e., comparing two sets of points) is a fundamental pattern matching problem. Thus many studies have been conducted in computational geometry [1,3,7,8] and pattern recognition [4,12].

Most of theoretical studies have focused on point matching under uniform transformations (e.g., translations, rigid motions and/or scalings) [1,3,7,8]. However, in some

---

<sup>☆</sup> Partially supported by a Grant-in-Aid “Genome Science” and a Grant-in-Aid No. 10780171 from The Ministry of Education, Science, Sports and Culture in Japan.

\* Corresponding author.

*E-mail addresses:* takutsu@ims.u-tokyo.ac.jp (Tatsuya Akutsu), kanaya@hydra.mki.co.jp (Kyotetsu Kanaya), akr@hydra.mki.co.jp (Akira Ohyama), afujiyam@lab.nig.ac.jp (Asao Fujiyama).

applications, non-uniform distortions may occur and thus pattern matching based on local similarity is important. Pattern matching of spots obtained by the two-dimensional (2-D) gel electrophoresis technique is an important example of such applications [2,6,13]. We have developed a system called DDGEL [9] for the analysis of 2-D gel electrophoresis images obtained from genomic DNA by means of the restriction landmark genomic scanning (RLGS) method [5]. Previous point matching methods are not directly applicable to this application because the positions of spots are distorted non-uniformly. Due to the recent progress in *proteomics*, analysis of 2-D gel electrophoresis images of proteins is an important issue. Therefore, a technique for matching of spots under non-uniform distortions is required [2,6].

On the other hand, in the field of pattern recognition, many heuristic algorithms have been developed for pattern matching under non-uniform distortions [2,4,6,12,13]. Appel et al., considered transformations based on second- and third-order polynomials to contend with non-uniform distortions appearing in electrophoresis image data [2]. However, their method (and many of other methods for electrophoresis image analysis) uses so-called *landmarks* in order to identify polynomials, where landmarks are spot pairs intensively marked in both images by the user and selected as putative matching pairs.

Several groups applied Delaunay graphs (Delaunay triangulations) and/or relative neighborhood graphs to the problem of point matching under non-uniform distortions [4,6,12,13]. In these methods, a Delaunay graph (or a relative neighborhood graph) is first computed from each set of points. A maximum common subgraph (or a similar structure) between two graphs is then computed. However, finding a maximum common subgraph is a time consuming process (it is NP-hard in general). Therefore, various heuristics were employed in the Delaunay-based approaches. It is reasonable to question whether or not such a time consuming procedure is essential for point matching under non-uniform distortions. This is the theoretical motivation behind our study.

This paper consists of two parts: a theoretical part and a practical part. In the theoretical part, we propose a simple definition for point matching under non-uniform distortions, where similar formalizations are given in [6,12]. We prove that the problem is NP-hard in two or more dimensions. This result answers the above question: time consuming search procedures such as finding a maximum common subgraph are essential for point matching under non-uniform distortions unless  $P=NP$ . On the other hand, we present a simple polynomial time algorithm for a special and one-dimensional case, which is similar to the well-known dynamic programming algorithms for approximate string matching and sequence alignment. In the practical part, we show a heuristic method for spot matching of 2-D electrophoresis gel images obtained from genomic DNA by means of the RLGS method [5]. Although this method is heuristic, it uses a variant of the dynamic programming algorithm mentioned above. The method is implemented in the DDGEL system and is tested using real gel image data.

The hardness result is interesting from a theoretical viewpoint because almost no NP-hardness results have been presented in the field of approximate point matching. The only exception was proposed by Iwanowski [8], who studied the problem of *approximate symmetry detection* in the plane and proved that it is NP-hard. It is interesting that reductions from the same problem (PLANAR 3SAT) were used both in this

paper and Iwanowski's paper. However, the reduction used in this paper is substantially different from that used by Iwanowski. Indeed, Iwanowski considered approximate symmetry under uniform distortions.

## 2. Definition of the problem

In this section, we propose a formal definition for point matching under non-uniform distortions. Although other appropriate definitions may have been suggested, this definition is simple and reasonable.

Let  $P = \{p_1, \dots, p_m\}$  and  $Q = \{q_1, \dots, q_n\}$  be point sets in  $d$ -dimensions, respectively. We call a set of pairs  $M = \{(p_{i_1}, q_{j_1}), \dots, (p_{i_l}, q_{j_l})\}$  a *match* if  $(\forall h \neq k)(p_{i_h} \neq p_{i_k} \text{ and } q_{j_h} \neq q_{j_k})$ .

**Definition 1** (Point matching under non-uniform distortions). Point matching under non-uniform distortions is, given a positive real  $\varepsilon$ , two point sets  $P = \{p_1, \dots, p_m\}$  and  $Q = \{q_1, \dots, q_n\}$  in  $d$ -dimensional Euclidean space, to find a maximum match  $M = \{(p_{i_1}, q_{j_1}), \dots, (p_{i_l}, q_{j_l})\}$  (i.e., a match  $M$  with the maximum cardinality) satisfying

$$(\forall k)(\forall h \neq k) \left( \frac{1}{1 + \varepsilon} < \frac{|q_{j_h} - q_{j_k}|}{|p_{i_h} - p_{i_k}|} < 1 + \varepsilon \right),$$

where  $|p - q|$  denotes the Euclidean distance between  $p$  and  $q$ .

It should be noted that  $P$  and  $Q$  can be interchanged because  $1/(1 + \varepsilon) < x/y < 1 + \varepsilon$  if and only if  $1/(1 + \varepsilon) < y/x < 1 + \varepsilon$ . This definition requires that local similarities between  $P$  and  $Q$  are preserved because the error for two point pairs must be small if the distances between points in the pairs are small.

In the above definition, global similarities are also taken into account to some extent. Of course, the problem might be defined such that only local similarities can be taken into account. However, in such a case, some threshold would be required to discriminate between local and global distances and thus the definition would be complex and less reasonable.

Here we show the difference between the proposed definition and  $\varepsilon$ -approximate congruence [1], where the latter is a typical point matching problem under uniform distortions and has been well studied in computational geometry. Recall that, in  $\varepsilon$ -approximate congruence,  $p_{i_1}$  matches  $q_{j_1}$  if and only if  $|q_{j_1} - p_{i_1}| < \varepsilon$ , where an adequate isometric transformation can be applied to  $P$  before computing the match.

In Fig. 1,  $P$  consists of four squares, each of which is of size  $1 \times 1$ . Consider the case of  $\varepsilon = 0.5$  in  $\varepsilon$ -approximate congruence. Then,  $P$  matches  $Q_1$  (i.e., there is a match  $M$  such that  $|M| = 16$ ), but does not match  $Q_2$ . On the other hand, in the case of  $\varepsilon = 0.5$  in point matching under non-uniform distortions,  $P$  does not match  $Q_1$  but matches  $Q_2$ . This example shows that local similarities are taken into account more in point matching under non-uniform distortions than in  $\varepsilon$ -approximate congruence.

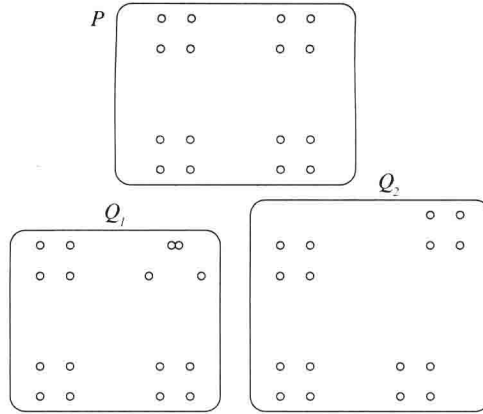


Fig. 1. This figure shows the difference between point matching under non-uniform distortions and point matching under uniform distortions.  $P$  matches  $Q_1$  but does not match  $Q_2$  under uniform distortions, whereas  $P$  does not match  $Q_1$  but matches  $Q_2$  under non-uniform distortions.

### 3. NP-hardness result for two-dimensional case

In Section 2, we proposed a definition for point matching under non-uniform distortions. However, this problem is NP-hard even for two-dimensions.

**Theorem 2.** *Point matching under non-uniform distortions in  $d$ -dimensions is NP-hard for any fixed number  $\varepsilon$  such that  $0 < \varepsilon < 0.44$  and for any fixed integer  $d \geq 2$ .*

**Proof.** We use a polynomial time reduction from PLANAR 3SAT [11]. Let  $C = \{c_1, c_2, \dots, c_N\}$  be an instance of PLANAR 3SAT over the set of variables  $V = \{v_1, v_2, \dots, v_K\}$ , where we assume that each clause  $c_i$  consists of 3 literals. Note that in PLANAR 3SAT, the graph  $G(V \cup C, E)$  must be planar, where  $E = \{\{v_i, c_j\} | v_i \in c_j \text{ or } \bar{v}_i \in c_j\} \cup \{\{v_i, v_{i+1}\}\} \cup \{\{v_1, v_K\}\}$ . Moreover, we assume that a grid embedding of  $G(V \cup C, E)$  is already obtained as shown in Fig. 2. A grid embedding of size  $O(N) \times O(N)$  can be computed in linear time from  $G(V \cup C, E)$  [10].

From the grid embedding of an instance of PLANAR 3SAT, we construct an instance  $(P, Q, \varepsilon)$  of the point matching problem. The instance is constructed such that the following holds: there exists a maximum match  $M$  satisfying  $|M| = |P|$  if and only if there exists a truth assignment satisfying all clauses. The construction will be formed from several components, which can be partitioned into three parts and grouped according to their intended function: *truth-setting* components, *satisfaction-testing* components, and *routing* components.

First, we will discuss the satisfaction-testing components (see also Fig. 3) since these form the core of the construction. For each clause  $c_i$ , we construct a set of

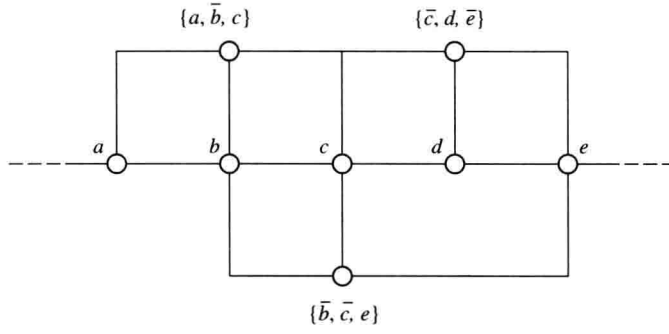


Fig. 2. Example of grid embedding of a planar graph for a 3SAT instance  $\{\{a, \bar{b}, c\}, \{\bar{c}, d, \bar{e}\}, \{\bar{b}, \bar{c}, e\}\}$ .

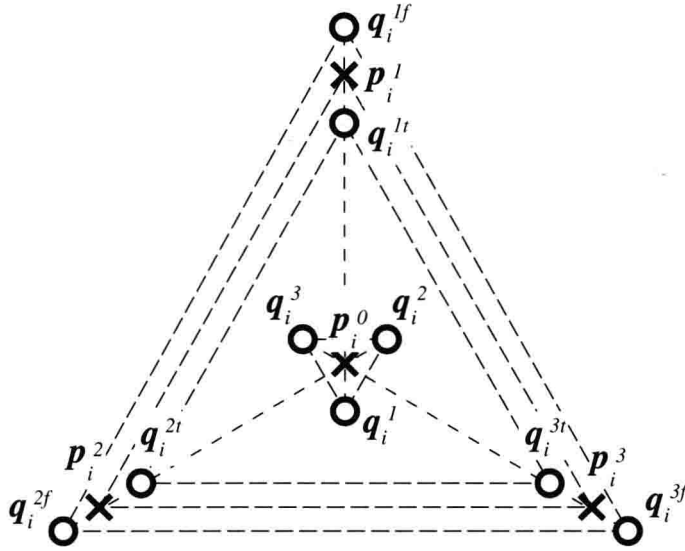


Fig. 3. Satisfaction-testing component.

points  $T_i = \{p_i^0, p_i^1, p_i^2, p_i^3, q_i^1, q_i^2, q_i^3, q_i^{1t}, q_i^{2t}, q_i^{3t}, q_i^{1f}, q_i^{2f}, q_i^{3f}\}$ , where

$$p_i^0 = (0, 0), \quad p_i^1 = (0, L), \quad p_i^2 = \left(-\frac{\sqrt{3}L}{2}, -\frac{L}{2}\right), \quad p_i^3 = \left(\frac{\sqrt{3}L}{2}, -\frac{L}{2}\right),$$

$$q_i^1 = (0, -\alpha L), \quad q_i^2 = \left(\frac{\sqrt{3}\alpha L}{2}, \frac{\alpha L}{2}\right), \quad q_i^3 = \left(-\frac{\sqrt{3}\alpha L}{2}, \frac{\alpha L}{2}\right),$$

$$q_i^{1t} = (0, (1 - \alpha)L), \quad q_i^{2t} = \left(-\frac{\sqrt{3}(1 - \alpha)L}{2}, -\frac{(1 - \alpha)L}{2}\right),$$

$$\begin{aligned} \mathbf{q}_i^{3t} &= \left( \frac{\sqrt{3}(1-\alpha)L}{2}, -\frac{(1-\alpha)L}{2} \right), \quad \mathbf{q}_i^{1f} = (0, (1+\alpha)L), \\ \mathbf{q}_i^{2f} &= \left( -\frac{\sqrt{3}(1+\alpha)L}{2}, -\frac{(1+\alpha)L}{2} \right), \quad \mathbf{q}_i^{3f} = \left( \frac{\sqrt{3}(1+\alpha)L}{2}, -\frac{(1+\alpha)L}{2} \right), \end{aligned}$$

and each  $T_i$  is to be translated to an appropriate position.  $L$  is any positive constant (for example, we can use  $L = 1$ ).

Here we let  $\varepsilon = 2\alpha$ . We can now see that the following relations hold for any  $\alpha$  such that  $0 < \alpha < 0.22$ :

$$\begin{aligned} (\forall j) \quad (|\mathbf{q}_i^{jf} - \mathbf{q}_i^j| &= (1+\varepsilon)L), \quad (\forall j) \quad \left( \left( \frac{1}{1+\varepsilon} \right) L < |\mathbf{q}_i^{jt} - \mathbf{q}_i^j| < (1+\varepsilon)L \right), \\ (\forall j) \quad (\forall k \neq j) \quad &\left( \left( \frac{1}{1+\varepsilon} \right) L < |\mathbf{q}_i^{jt} - \mathbf{q}_i^k| < (1+\varepsilon)L \right), \\ (\forall j) \quad (\forall k \neq j) \quad &\left( \left( \frac{1}{1+\varepsilon} \right) L < |\mathbf{q}_i^{jf} - \mathbf{q}_i^k| < (1+\varepsilon)L \right). \end{aligned}$$

It should be noted that  $(1/(1+\varepsilon))L < |\mathbf{q}_i^{jt} - \mathbf{q}_i^k|$  ( $k \neq j$ ) does not hold if  $\alpha > 0.23$ .

Because of the above relationships, the following must hold in order to satisfy  $|M| = |P|$ :

- (a) if  $(\mathbf{p}_i^0, \mathbf{q}_i^1) \in M$ , then  $(\mathbf{p}_i^1, \mathbf{q}_i^{1t}) \in M$ ,  $(\mathbf{p}_i^2, \mathbf{q}_i^{2t}) \in M$  or  $(\mathbf{p}_i^2, \mathbf{q}_i^{2f}) \in M$ , and,  $(\mathbf{p}_i^3, \mathbf{q}_i^{3t}) \in M$  or  $(\mathbf{p}_i^3, \mathbf{q}_i^{3f}) \in M$ ,
- (b) if  $(\mathbf{p}_i^0, \mathbf{q}_i^2) \in M$ , then  $(\mathbf{p}_i^2, \mathbf{q}_i^{2t}) \in M$ ,  $(\mathbf{p}_i^1, \mathbf{q}_i^{1t}) \in M$  or  $(\mathbf{p}_i^1, \mathbf{q}_i^{1f}) \in M$ , and,  $(\mathbf{p}_i^3, \mathbf{q}_i^{3t}) \in M$  or  $(\mathbf{p}_i^3, \mathbf{q}_i^{3f}) \in M$ ,
- (c) if  $(\mathbf{p}_i^0, \mathbf{q}_i^3) \in M$ , then  $(\mathbf{p}_i^3, \mathbf{q}_i^{3t}) \in M$ ,  $(\mathbf{p}_i^1, \mathbf{q}_i^{1t}) \in M$  or  $(\mathbf{p}_i^1, \mathbf{q}_i^{1f}) \in M$ , and,  $(\mathbf{p}_i^2, \mathbf{q}_i^{2t}) \in M$  or  $(\mathbf{p}_i^2, \mathbf{q}_i^{2f}) \in M$ .

From this, we can see that at least one of  $(\mathbf{p}_i^1, \mathbf{q}_i^{1t}) \in M$ ,  $(\mathbf{p}_i^2, \mathbf{q}_i^{2t}) \in M$  and  $(\mathbf{p}_i^3, \mathbf{q}_i^{3t}) \in M$  must hold in order to satisfy  $|M| = |P|$ . Here we assume  $c_i = \{v_1, v_2, v_3\}$ . Then, case (a) corresponds to the case where  $v_1$  is satisfied, case (b) corresponds to the case where  $v_2$  is satisfied, and case (c) corresponds to the case where  $v_3$  is satisfied. Thus, at least one of  $(\mathbf{p}_i^1, \mathbf{q}_i^{1t}) \in M$ ,  $(\mathbf{p}_i^2, \mathbf{q}_i^{2t}) \in M$  and  $(\mathbf{p}_i^3, \mathbf{q}_i^{3t}) \in M$  must hold in order to satisfy clause  $c_i$ .

Next we will discuss the truth-setting components. For each variable  $v_i$ , we construct a truth-setting component (see Fig. 4). Let  $(x_i, 0)$  be the position for variable  $v_i$  in the grid embedding of  $G(V \cup C, E)$ . We construct the point sets  $P_i$ ,  $Q_i^t$  and  $Q_i^f$  by

$$\begin{aligned} P_i &= \{(x_i, 0)\} \cup \left\{ \left( x_i, \frac{D}{2} \right), \left( x_i, -\frac{D}{2} \right) \right\} \\ &\quad \cup \left\{ \left( x_i, \frac{D}{2} + jD \right), \left( x_i, -\frac{D}{2} - jD \right) \mid j = 1, \dots \right\}, \\ Q_i^t &= \{(x_i - D, 0)\} \cup \left\{ \left( x_i - D, \frac{D}{2} \right), \left( x_i - D, -\frac{D}{2} \right) \right\} \end{aligned}$$



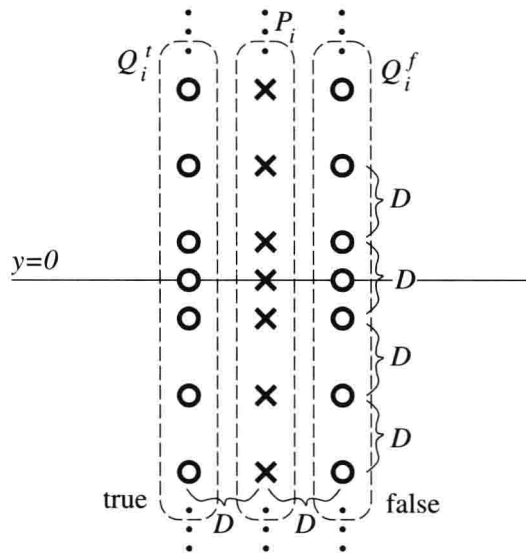


Fig. 4. Truth-setting component.

$$\cup \left\{ \left( x_i - D, \frac{D}{2} + jD \right), \left( x_i - D, -\frac{D}{2} - jD \right) \mid j = 1, \dots \right\},$$

$$Q_i^f = \{(x_i + D, 0)\} \cup \left\{ \left( x_i + D, \frac{D}{2} \right), \left( x_i + D, -\frac{D}{2} \right) \right\}$$

$$\cup \left\{ \left( x_i + D, \frac{D}{2} + jD \right), \left( x_i + D, -\frac{D}{2} - jD \right) \mid j = 1, \dots \right\},$$

where  $D = \alpha L$ .

For these point sets, the following property holds for any  $\varepsilon$  such that  $0 < \varepsilon < 0.44$ : either  $P_i$  matches  $Q_i^t$  or  $P_i$  matches  $Q_i^f$  under the condition that every point in  $P_i$  matches a point in  $Q$ . These point sets are constructed such that “ $P_i$  matches  $Q_i^t$ ” corresponds to “ $v_i = \text{true}$ ” and “ $P_i$  matches  $Q_i^f$ ” corresponds to “ $v_i = \text{false}$ ”.

Next we will discuss the routing components. According to the grid embedding of  $G(V \cup C, E)$ , the truth-setting components are connected to the satisfaction-testing components. In order to do this, the following types of *gadgets* are used:

- (i) copying a truth assignment,
- (ii) inverting a truth assignment on  $v_i$  (i.e., creating  $\bar{v}_i$ ),
- (iii) connecting a truth assignment to a satisfaction-testing component.

Gadgets of type (i) are constructed in the following way. Each gadget has the form shown in Fig. 5 and consists of three point sets:  $\hat{P}$ ,  $\hat{Q}^t$  and  $\hat{Q}^f$ . In Fig. 5,  $\hat{Q}^t$  consists of the points on the bold lines,  $\hat{Q}^f$  consists of the points on the dashed bold lines and  $\hat{P}$  consists of the points on the thin lines. Explicit coordinates of the points are given