Franco P. Preparata
Qizhi Fang (Eds.)

LNCS 4613

Frontiers in Algorithmics

First Annual International Workshop, FAW 2007
Lanzhou, China, August 2007
Proceedings

Springer

Franco P. Preparata   Qizhi Fang (Eds.)

Frontiers in Algorithmics

First Annual International Workshop, FAW 2007
Lanzhou, China, August 1-3, 2007
Proceedings

Springer

Volume Editors

Franco P. Preparata
Brown University, Department of Computer Science
115 Waterman St., Providence, RI 02912-1910, USA
E-mail: franco@cs.brown.edu

Qizhi Fang
Ocean University of China, Department of Mathematics
23 Hong Kong Eastern Road, Qingdao 266071, China
E-mail: qfang@ouc.edu.cn

Library of Congress Control Number: 2007931053

CR Subject Classification (1998): F.2.2, G.2.1-2, G.1.2, G.1.6, C.2.2, I.3.5, E.1

LNCS Sublibrary: SL 1 – Theoretical Computer Science and General Issues

ISSN 0302-9743
ISBN-10 3-540-73813-4 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-73813-8 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media

springer.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12097180 06/3180 5 4 3 2 1 0

Lecture Notes in Computer Science 4613

Lecture Notes in Computer Science

For information about Vols. 1–4525

please contact your bookseller or Springer

Vol. 4613: F.P. Preparata, Q. Fang (Eds.), Frontiers in Algorithmics. XI, 348 pages. 2007.

Vol. 4612: I. Miguel, W. Ruml (Eds.), Abstraction, Reformulation, and Approximation. XI, 418 pages. 2007. (Sublibrary LNAI).

Vol. 4611: J. Indulska, J. Ma, L.T. Yang, T. Ungerer, J. Cao (Eds.), Ubiquitous Intelligence and Computing. XXIII, 1257 pages. 2007.

Vol. 4610: B. Xiao, L.T. Yang, J. Ma, C. Muller-Schloer, Y. Hua (Eds.), Autonomic and Trusted Computing. XVIII, 571 pages. 2007.

Vol. 4608: H.W. Schmidt, I. Crnkovic, G.T. Heineman, J.A. Stafford (Eds.), Component-Based Software Engineering. XII, 283 pages. 2007.

Vol. 4607: L. Baresi, P. Fraternali, G.-J. Houben (Eds.), Web Engineering. XVI, 576 pages. 2007.

Vol. 4606: A. Pras, M. van Sinderen (Eds.), Dependable and Adaptable Networks and Services. XIV, 149 pages. 2007.

Vol. 4605: D. Papadias, D. Zhang, G. Kollios (Eds.), Advances in Spatial and Temporal Databases. X, 479 pages. 2007.

Vol. 4604: U. Priss, S. Polovina, R. Hill (Eds.), Conceptual Structures: Knowledge Architecture for Smart Applications. XII, 514 pages. 2007. (Sublibrary LNAI).

Vol. 4603: F. Pfenning (Ed.), Automated Deduction – CADE-21. XII, 522 pages. 2007. (Sublibrary LNAI).

Vol. 4602: S. Barker, G.-J. Ahn (Eds.), Data and Applications Security XXI. X, 291 pages. 2007.

Vol. 4600: H. Comon-Lundh, C. Kirchner, H. Kirchner, Rewriting, Computation and Proof. XVI, 273 pages. 2007.

Vol. 4599: S. Vassiliadis, M. Berekovic, T.D. Hämäläinen (Eds.), Embedded Computer Systems: Architectures, Modeling, and Simulation. XVIII, 466 pages. 2007.

Vol. 4598: G. Lin (Ed.), Computing and Combinatorics. XII, 570 pages. 2007.

Vol. 4597: P. Perner (Ed.), Advances in Data Mining. XI, 353 pages. 2007. (Sublibrary LNAI).

Vol. 4596: L. Arge, C. Cachin, T. Jurdziński, A. Tarlecki (Eds.), Automata, Languages and Programming. XVII, 953 pages. 2007.

Vol. 4595: D. Bošnački, S. Edelkamp (Eds.), Model Checking Software. X, 285 pages. 2007.

Vol. 4594: R. Bellazzi, A. Abu-Hanna, J. Hunter (Eds.), Artificial Intelligence in Medicine. XVI, 509 pages. 2007. (Sublibrary LNAI).

Vol. 4592: Z. Kedad, N. Lammari, E. Métais, F. Meziane, Y. Rezgui (Eds.), Natural Language Processing and Information Systems. XIV, 442 pages. 2007.

Vol. 4591: J. Davies, J. Gibbons (Eds.), Integrated Formal Methods. IX, 660 pages. 2007.

Vol. 4590: W. Damm, H. Hermanns (Eds.), Computer Aided Verification. XV, 562 pages. 2007.

Vol. 4589: J. Münch, P. Abrahamsson (Eds.), Product-Focused Software Process Improvement. XII, 414 pages. 2007.

Vol. 4588: T. Harju, J. Karhumäki, A. Lepistö (Eds.), Developments in Language Theory. XI, 423 pages. 2007.

Vol. 4587: R. Cooper, J. Kennedy (Eds.), Data Management. XIII, 259 pages. 2007.

Vol. 4586: J. Pieprzyk, H. Ghodosi, E. Dawson (Eds.), Information Security and Privacy. XIV, 476 pages. 2007.

Vol. 4585: M. Kryszkiewicz, J.F. Peters, H. Rybinski, A. Skowron (Eds.), Rough Sets and Intelligent Systems Paradigms. XIX, 836 pages. 2007. (Sublibrary LNAI).

Vol. 4584: N. Karssemeijer, B. Lelieveldt (Eds.), Information Processing in Medical Imaging. XX, 777 pages. 2007.

Vol. 4583: S.R. Della Rocca (Ed.), Typed Lambda Calculi and Applications. X, 397 pages. 2007.

Vol. 4582: J. Lopez, P. Samarati, J.L. Ferrer (Eds.), Public Key Infrastructure. XI, 375 pages. 2007.

Vol. 4581: A. Petrenko, M. Veanes, J. Tretmans, W. Grieskamp (Eds.), Testing of Software and Communicating Systems. XII, 379 pages. 2007.

Vol. 4580: B. Ma, K. Zhang (Eds.), Combinatorial Pattern Matching. XII, 366 pages. 2007.

Vol. 4579: B. M. Hämmerli, R. Sommer (Eds.), Detection of Intrusions and Malware, and Vulnerability Assessment. X, 251 pages. 2007.

Vol. 4578: F. Masulli, S. Mitra, G. Pasi (Eds.), Applications of Fuzzy Sets Theory. XVIII, 693 pages. 2007. (Sublibrary LNAI).

Vol. 4577: N. Sebe, Y. Liu, Y.-t. Zhuang (Eds.), Multimedia Content Analysis and Mining. XIII, 513 pages. 2007.

Vol. 4576: D. Leivant, R. de Queiroz (Eds.), Logic, Language, Information and Computation. X, 363 pages. 2007.

Vol. 4575: T. Takagi, T. Okamoto, E. Okamoto, T. Okamoto (Eds.), Pairing-Based Cryptography – Pairing 2007. XI, 408 pages. 2007.

Vol. 4574: J. Derrick, J. Vain (Eds.), Formal Techniques for Networked and Distributed Systems – FORTE 2007. XI, 375 pages. 2007.

Vol. 4573: M. Kauers, M. Kerber, R. Miner, W. Windsteiger (Eds.), Towards Mechanized Mathematical Assistants. XIII, 407 pages. 2007. (Sublibrary LNAI).

Vol. 4572: F. Stajano, C. Meadows, S. Capkun, T. Moore (Eds.), Security and Privacy in Ad-hoc and Sensor Networks. X, 247 pages. 2007.

Vol. 4571: P. Perner (Ed.), Machine Learning and Data Mining in Pattern Recognition. XIV, 913 pages. 2007. (Sublibrary LNAI).

Vol. 4570: H.G. Okuno, M. Ali (Eds.), New Trends in Applied Artificial Intelligence. XXI, 1194 pages. 2007. (Sublibrary LNAI).

Vol. 4569: A. Butz, B. Fisher, A. Krüger, P. Olivier, S. Owada (Eds.), Smart Graphics. IX, 237 pages. 2007.

Vol. 4566: M.J. Dainoff (Ed.), Ergonomics and Health Aspects of Work with Computers. XVIII, 390 pages. 2007.

Vol. 4565: D.D. Schmorrow, L.M. Reeves (Eds.), Foundations of Augmented Cognition. XIX, 450 pages. 2007. (Sublibrary LNAI).

Vol. 4564: D. Schuler (Ed.), Online Communities and Social Computing. XVII, 520 pages. 2007.

Vol. 4563: R. Shumaker (Ed.), Virtual Reality. XXII, 762 pages. 2007.

Vol. 4562: D. Harris (Ed.), Engineering Psychology and Cognitive Ergonomics. XXIII, 879 pages. 2007. (Sublibrary LNAI).

Vol. 4561: V.G. Duffy (Ed.), Digital Human Modeling. XXIII, 1068 pages. 2007.

Vol. 4560: N. Aykin (Ed.), Usability and Internationalization, Part II. XVIII, 576 pages. 2007.

Vol. 4559: N. Aykin (Ed.), Usability and Internationalization, Part I. XVIII, 661 pages. 2007.

Vol. 4558: M.J. Smith, G. Salvendy (Eds.), Human Interface and the Management of Information, Part II. XXIII, 1162 pages. 2007.

Vol. 4557: M.J. Smith, G. Salvendy (Eds.), Human Interface and the Management of Information, Part I. XXII, 1030 pages. 2007.

Vol. 4556: C. Stephanidis (Ed.), Universal Access in Human-Computer Interaction, Part III. XXII, 1020 pages. 2007.

Vol. 4555: C. Stephanidis (Ed.), Universal Access in Human-Computer Interaction, Part II. XXII, 1066 pages. 2007.

Vol. 4554: C. Stephanidis (Ed.), Universal Acess in Human Computer Interaction, Part I. XXII, 1054 pages. 2007.

Vol. 4553: J.A. Jacko (Ed.), Human-Computer Interaction, Part IV. XXIV, 1225 pages. 2007.

Vol. 4552: J.A. Jacko (Ed.), Human-Computer Interaction, Part III. XXI, 1038 pages. 2007.

Vol. 4551: J.A. Jacko (Ed.), Human-Computer Interaction, Part II. XXIII, 1253 pages. 2007.

Vol. 4550: J.A. Jacko (Ed.), Human-Computer Interaction, Part I. XXIII, 1240 pages. 2007.

Vol. 4549: J. Aspnes, C. Scheideler, A. Arora, S. Madden (Eds.), Distributed Computing in Sensor Systems. XIII, 417 pages. 2007.

Vol. 4548: N. Olivetti (Ed.), Automated Reasoning with Analytic Tableaux and Related Methods. X, 245 pages. 2007. (Sublibrary LNAI).

Vol. 4547: C. Carlet, B. Sunar (Eds.), Arithmetic of Finite Fields. XI, 355 pages. 2007.

Vol. 4546: J. Kleijn, A. Yakovlev (Eds.), Petri Nets and Other Models of Concurrency – ICATPN 2007. XI, 515 pages. 2007.

Vol. 4545: H. Anai, K. Horimoto, T. Kutsia (Eds.), Algebraic Biology. XIII, 379 pages. 2007.

Vol. 4544: S. Cohen-Boulakia, V. Tannen (Eds.), Data Integration in the Life Sciences. XI, 282 pages. 2007. (Sublibrary LNBI).

Vol. 4543: A.K. Bandara, M. Burgess (Eds.), Inter-Domain Management. XII, 237 pages. 2007.

Vol. 4542: P. Sawyer, B. Paech, P. Heymans (Eds.), Requirements Engineering: Foundation for Software Quality. IX, 384 pages. 2007.

Vol. 4541: T. Okadome, T. Yamazaki, M. Makhtari (Eds.), Pervasive Computing for Quality of Life Enhancement. IX, 248 pages. 2007.

Vol. 4539: N.H. Bshouty, C. Gentile (Eds.), Learning Theory. XII, 634 pages. 2007. (Sublibrary LNAI).

Vol. 4538: F. Escolano, M. Vento (Eds.), Graph-Based Representations in Pattern Recognition. XII, 416 pages. 2007.

Vol. 4537: K.C.-C. Chang, W. Wang, L. Chen, C.A. Ellis, C.-H. Hsu, A.C. Tsoi, H. Wang (Eds.), Advances in Web and Network Technologies, and Information Management. XXIII, 707 pages. 2007.

Vol. 4536: G. Concas, E. Damiani, M. Scotto, G. Succi (Eds.), Agile Processes in Software Engineering and Extreme Programming. XV, 276 pages. 2007.

Vol. 4534: I. Tomkos, F. Neri, J. Solé Pareta, X. Masip Bruin, S. Sánchez Lopez (Eds.), Optical Network Design and Modeling. XI, 460 pages. 2007.

Vol. 4533: F. Baader (Ed.), Term Rewriting and Applications. XII, 419 pages. 2007.

Vol. 4531: J. Indulska, K. Raymond (Eds.), Distributed Applications and Interoperable Systems. XI, 337 pages. 2007.

Vol. 4530: D.H. Akehurst, R. Vogel, R.F. Paige (Eds.), Model Driven Architecture- Foundations and Applications. X, 219 pages. 2007.

Vol. 4529: P. Melin, O. Castillo, L.T. Aguilar, J. Kacprzyk, W. Pedrycz (Eds.), Foundations of Fuzzy Logic and Soft Computing. XIX, 830 pages. 2007. (Sublibrary LNAI).

Vol. 4528: J. Mira, J.R. Álvarez (Eds.), Nature Inspired Problem-Solving Methods in Knowledge Engineering, Part II. XXII, 650 pages. 2007.

Vol. 4527: J. Mira, J.R. Álvarez (Eds.), Bio-inspired Modeling of Cognitive Tasks, Part I. XXII, 630 pages. 2007.

Vol. 4526: M. Malek, M. Reitenspieß, A. van Moorsel (Eds.), Service Availability. X, 155 pages. 2007.

Preface

FAW 2007, the 1st International "Frontiers in Algorithmics Workshop" took place in Lanzhou, China, August 1–3, 2007. The FAW symposium aims to provide a focused forum on current trends in research on algorithms, discrete structures, and their applications, and to bring together international experts at the research frontiers in those areas so as to exchange ideas and to present significant new results. In response to the Call for Papers, a total of 141 papers were submitted from 16 countries and regions, of which 33 were accepted. These papers were selected for nine special focus tracks in the areas of bioinformatics, discrete structures, geometric information processing and communication, games and incentive analysis, graph algorithms, Internet algorithms and protocols, parameterized algorithms, design and analysis of heuristics, approximate and online algorithms, and algorithms in medical applications.

We would like to thank the Conference General Chair, Maocheng Cai and Hao Li, and Advising Committee Chair, Danny Chen, for their leadership, advice and help on crucial matters concerning the conference. We would like to thank the International Program Committee and the external referees for spending their valuable time and effort in the review process. It was a wonderful experience to work with them.

Finally, we would like to thank the Organizing Committee, led by Lian Li and Xiaotie Deng, for their contribution to making this conference a success. We would also like to thank our sponsors, the ICCM Laboratory of Lanzhou University, for kindly offering the financial and clerical support that made the conference possible and enjoyable.

August 2007

Franco Preparata
Qizhi Fang

Organization

FAW 2007 was sponsored by the ICCM Laboratory of Lanzhou University, China.

Executive Committee

General Chairs	Mao-cheng Cai (Chinese Academy of Science, China)
	Hao Li (Laboratoire de Recherche en Informatique, France)
Program Co-chairs	Franco P. Preparata (Brown University, USA)
	Qizhi Fang (Ocean University of China, China)
Organization Chair	Lian Li (Lanzhou University, China)
Organization Co-chair	Xiaotie Deng (City University of Hong Kong, China)
Advising Committee Chair	Danny Chen (University of Notre Dame, USA)

Program Committee

Helmut Alt (Freie Universität Berlin)
Cristina Bazgan (Université Paris Dauphine)
Stephan Brandt (Technische Universität Ilmenau)
Hajo Broersma (University of Durham)
Leizhen Cai (The Chinese University of Hong Kong)
Bernard M. Chazelle (Princeton University)
Jianer Chen (Texas A&M University)
Francis Y.L. Chin (The University of Hong Kong)
Genghua Fan (Fuzhou University)
Evelyne Flandrin (Université Paris-Sud)
Anna Gal (University of Texas at Austin)
Ronald L. Graham (University of Californi at San Diego)
Toshihide Ibaraki (Kwansei Gakuin University)
Hiroshi Imai (University of Tokyo)
Kazuo Iwama (Kyoto University)
Kazuo Iwano (IBM Japan, Software Development Laboratory)
Ming-Yang Kao (Northwestern University)
Jean-Claude Latombe (Stanford University)
Der-Tsai Lee (Institute of Information Science Academia Sinica)
Lian Li (Lanzhou University)

Detlef Seese (Universität Karlsruhe (TH))
Yaoyun Shi (University of Michigan)
Lusheng Wang (City University of Hong Kong)
Roger Yu (Thompson Rivers University, also Nankai University)
Louxin Zhang (National University of Singapore)

External Reviewers

Juergen Banke
Stephan Chalup
Tobias Dietrich
Bin Fu
Gregory Gutin
Horace Ip
Yuval Ishai
Ioannis Ivrissimtzis
Iyad Kanj
Gyula Y. Katona
Koji Kojima
Arie Koster
Cheng-Chung Li
Chung-Shou Liao
Tien-Ching Lin
Ching-Chi Lin
Bin Ma
Bert Marchal
Joachim Melcher
Gerhard Post
Artem Pyatkin
Amir Safari
Ludmila Scharf
Marc Scherfenberg
Ingo Schiermeyer
Sven Scholz
Xiaoxun Sun
Helen Tu
Aaron Wong
Mingyu Xiao
Guomin Yang
Teng-Kai Yu
Yong Zhang
Shenggui Zhang
Magnus Bordewich

Table of Contents

Geometric Algorithms for the Constrained 1-D K-Means Clustering Problems and IMRT Applications*

Danny Z. Chen, Mark A. Healy, Chao Wang, and Bin Xu**

Department of Computer Science and Engineering
University of Notre Dame, Notre Dame, IN 46556, USA
{chen,mhealy4,cwang1,bxu}@cse.nd.edu

Abstract. In this paper, we present efficient geometric algorithms for the discrete constrained 1-D K-means clustering problem and extend our solutions to the continuous version of the problem. One key clustering constraint we consider is that the maximum difference in each cluster cannot be larger than a given threshold. These constrained 1-D K-means clustering problems appear in various applications, especially in intensity-modulated radiation therapy (IMRT). Our algorithms improve the efficiency and accuracy of the heuristic approaches used in clinical IMRT treatment planning.

Keywords: K-means clustering, staircase-Monge property, matrix search algorithm, minimum-weight K-link path algorithm, intensity modulated radiation therapy (IMRT).

1 Introduction

Data clustering is a fundamental problem that arises in many applications (e.g., data mining, information retrieval, pattern recognition, biomedical informatics, and statistics). The main objective of data clustering is to partition a given data set into clusters (i.e., subsets) based on certain optimization criteria and subject to certain clustering constraints.

In this paper, we consider the discrete and continuous constrained 1-D K-means clustering problems and their applications in intensity-modulated radiation therapy (IMRT). The definitions of these two data clustering problems are given as follows.

The **discrete constrained 1-D K-means clustering problem**: We are given a positive bandwidth parameter $\delta \in \mathbb{R}$, integers K and n with $1 < K < n$, n real numbers $x_1, x_2, \ldots, x_n$ with $x_1 < x_2 < \cdots < x_n$, and a positive real-valued probability function $P : \{x_1, x_2, \ldots x_n\} \rightarrow (0, 1)$ such that $\sum_{i=1}^{n} P(x_i) = 1$. For any j and l with $0 \leq l < j \leq n$, we define

$$\mu[l, j] = \frac{\sum_{i=l+1}^{j} P(x_i) * x_i}{\sum_{i=l+1}^{j} P(x_i)}$$

* This research was supported in part by NSF Grant CCF-0515203.
** Corresponding author.

F.P. Preparata and Q. Fang (Eds.): FAW 2007, LNCS 4613, pp. 1–13, 2007.

and

$$V[l,j] = \begin{cases} +\infty, & \text{when } x_j - x_{l+1} > \delta \\ \sum_{i=l+1}^{j} P(x_i)(x_i - \mu[l,j])^2, & \text{when } x_j - x_{l+1} \le \delta \end{cases}$$

We seek a sequence $q = (q_1, q_2, \ldots, q_{K-1})$ of $K-1$ integers with $0 < q_1 < q_2 < \cdots < q_{K-1} < n$ (for convenience, $q_0 \triangleq 0$ and $q_K \triangleq n$) such that the total error $E(q) = \sum_{k=1}^{K} V[q_{k-1}, q_k]$ is minimized. This discrete constrained problem arises in IMRT applications [4].

The **continuous constrained 1-D K-means clustering problem:** We are given a positive bandwidth parameter $\delta \in \mathbb{R}$, integers K and n with $1 < K < n$, and a positive real-valued (density) function $f : [x_b, x_e] \rightarrow [0,1]$, where x_b and x_e ($x_b < x_e$) are real numbers and $\int_{x_b}^{x_e} f(x)dx = 1$. For any values α and β with $x_b \le \alpha < \beta \le x_e$, we define

$$\tilde{\mu}[\alpha,\beta] = \frac{\int_\alpha^\beta f(x) * x dx}{\int_\alpha^\beta f(x) dx}$$

and

$$\tilde{V}[\alpha,\beta] = \begin{cases} +\infty, & \text{when } \beta - \alpha > \delta \\ \int_\alpha^\beta f(x)(x - \tilde{\mu}[\alpha,\beta])^2 dx, & \text{when } \beta - \alpha \le \delta \end{cases}$$

We seek a sequence $\theta = (\theta_1, \theta_2, \ldots, \theta_{K-1})$ of $K-1$ real numbers with $x_b < \theta_1 < \theta_2 < \cdots < \theta_{K-1} < x_e$ (for convenience, $\theta_0 \triangleq x_b$ and $\theta_K \triangleq x_e$), such that the total error $\tilde{E}(\theta) = \sum_{j=1}^{K} \tilde{V}[\theta_{j-1}, \theta_j]$ is minimized.

Note that in the above definitions, $V[q_{j-1}, q_j] = +\infty$ when $x_{q_j} - x_{q_{j-1}+1} > \delta$, and $\tilde{V}[\theta_{j-1}, \theta_j] = +\infty$ when $\theta_j - \theta_{j-1} > \delta$. Thus both the problems actually have a common constraint, i.e., the maximum difference in any cluster cannot be greater than the bandwidth parameter δ.

Algorithms for these two problems without the above constraint have been widely used in many areas such as signal processing, data compression, and information theory. Various techniques have been applied to solve the unconstrained versions, such as quantization [8,17], K-means clustering [9], and computational geometry [1,2,3,7,16]. However, the constrained versions are also important to some applications such as IMRT, which motivate our study.

Efficient algorithms for the discrete *unconstrained* 1-D K-means clustering problem have been known. Wu [17] modeled the optimal quantization problem (a variation of the unconstrained version) as a K-link shortest path problem and gave an $O(Kn)$ time algorithm based on dynamic programming and Monge matrix search techniques [1,2]. Aggarwal *et al.* [3] showed that the K-link shortest path problem on directed acyclic graphs (DAG) that satisfy the Monge property [15,1,2] can be solved in $O(n\sqrt{K \log n})$ time by using a refined parametric search paradigm, and Schieber [16] improved the time bound to $O(n 2^{\sqrt{\log K \log \log n}})$. All these three algorithms exploit the Monge property to find an optimal solution efficiently. A related work due to Hassin and Tamir [10] formulated a class of location problems using a general facility location model; their solution for the

p-median problem is also based on dynamic programming and matrix search techniques.

The continuous *unconstrained* 1-D K-means problem was studied independently by Lloyd [11] and Max [13], who gave algorithms based on iterative numerical methods. The convergence speed of their algorithms was improved in [12]. These algorithms, however, are able to find only a local minimal solution instead of a global minimum. Wu [17] showed that by discretizing the continuous input function f, a computationally feasible global search algorithm could be obtained.

The constrained 1-D K-means clustering problems appear in the radiation dose calculation process of intensity-modulated radiation therapy (IMRT) for cancer treatment [4,14,18]. In IMRT treatment planning, a radiation dose prescription (i.e., a dose function) is first computed for a cancer patient by a treatment planning system. The initially computed dose function is in either a discrete form (i.e., a piecewise linear function) or a continuous form (e.g., a certain continuous smooth function), which is often too complicated to be deliverable. Hence, the initial dose function needs to be processed or simplified into a deliverable form called *intensity profile*. The resulting intensity profile should be as close as possible to the initial dose function both locally and globally to minimize error in the treatment plan. The constrained 1-D K-means clustering problems model the problem of computing an intensity profile from an initial dose function [4,14,18], in which the bandwidth parameter δ specifies the allowed local deviation error and the number K of clusters indicates the (delivery) complexity of the resulting intensity profile (e.g., see Section 4 and Figure 2).

Several algorithms for the constrained 1-D K-means clustering problems have been given in medical literature and used in clinical IMRT planning systems [4,18]. However, these algorithms use only heuristic methods to determine the clusters iteratively [4,14,18], and can be trapped in a local minimal (as shown by our experimental results in Section 5). Also, no theoretical analysis was given for their convergence speed. The clustering constraint defined at the beginning of this section is often used [4]. But, in certain medical settings, another clustering constraint (e.g., in [18]; see Section 2.5) that is quite similar to yet slightly more restricted than the constraint defined in this section is also used. As we will show later, such seemingly slight constraint variations can have quite different computational and algorithmic implications.

Our results in this paper are summarized as follows.

1. We present efficient geometric algorithms for computing optimal solutions to the discrete constrained 1-D K-means problem that is defined in this section. Depending on the relative values of n and K, our algorithms run in $O(Kn)$ or $O(n2^{\sqrt{\log K \log\log n}})$ time, by exploiting the Monge property [15,1,2,16,17] of the problem (see Section 2).
2. We also consider a similar yet slightly more restricted clustering constraint [18] (to be defined and discussed in Section 2.5), and show that the Monge property does not hold for this constraint variation. Our algorithm for this discrete constrained problem version takes $O(K(n + |E'|))$ time, where $|E'|$

is the number of edges in a graph that models the problem ($|E'| \leq n^2$, but in practice $|E'|$ can be much smaller than $O(n^2)$).
3. We extend our solutions to the continuous constrained 1-D K-means problem, by transforming the continuous case to the discrete case (see Section 3).
4. We show that our constrained 1-D K-means algorithms are useful in IMRT applications (see Section 4), and give some experimental results on comparing our solutions with those computed by the heuristic methods in medical literature [4,14,18] (see Section 5).

2 Algorithms for the Discrete Constrained 1-D K-Means Problem

This section presents our algorithms for the discrete constrained 1-D K-means clustering problem.

2.1 Computing the Minimum Number K of Clusters

The number K of clusters can be an input value. However, in some applications such as IMRT, K needs to be as small as possible and thus needs to be computed. Wu *et al.* [18] gave a greedy algorithm for finding the minimum cluster number K, as follows. The input values $x_1, x_2, \ldots, x_n$ are scanned in ascending order and partitioned into K groups:

$$(x_1, \ldots, x_{q_1}), (x_{q_1+1}, \ldots, x_{q_2}), \ldots, (x_{q_{K-1}+1}, \ldots, x_n)$$

such that for any k with $1 \leq k \leq K$, $x_{q_k} - x_{q_{k-1}+1} \leq \delta$ and $x_{q_k+1} - x_{q_{k-1}+1} > \delta$. Clearly, K can be computed in $O(n)$ time.

2.2 Reformulation of the Discrete Constrained 1-D K-Means Problem

In this section, we model the discrete constrained 1-D K-means clustering problem as a K-link shortest path problem on a directed acyclic graph (DAG) $G = (U, E)$, which is defined as follows (see Figure 1). The vertex set $U = \{u_0, u_1, u_2, \ldots, u_n\}$. For any two vertices u_l and u_j ($l < j$), we put an edge in G from u_l to u_j with a weight $V[l, j]$.

Clearly, G is a complete DAG with edge weights. Any K-link path from u_0 to u_n in G, say $p = u_0 \to u_{q_1} \to u_{q_2} \to \cdots \to u_{q_{K-1}} \to u_n$, corresponds to a feasible solution $q = (q_1, q_2, \ldots, q_{K-1})$ for the discrete constrained 1-D K-means clustering problem, and *vice versa*. For any path in G, define its *weight* as the sum of the weights of all its edges. It is easy to see that an optimal solution for the discrete constrained 1-D K-means clustering problem corresponds to a shortest K-link path from u_0 to u_n in the DAG G.

The DAG G thus defined has $O(n)$ vertices and $O(n^2)$ edges. In Section 2.3, we will show that the weight of any edge in G can be computed in $O(1)$ time after an $O(n)$ time preprocess. Hence, the graph G can be represented *implicitly*, that is, after the $O(n)$ time preprocess, any vertex and edge of G can be obtained in $O(1)$

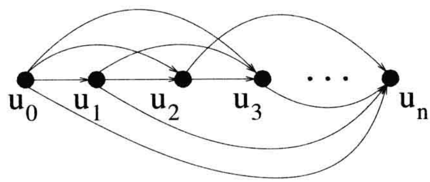

Fig. 1. A weighted, complete directed acyclic graph (DAG)

time. Thus, a K-link shortest path in G can be computed in $O(K(|U|+|E|)) = O(Kn^2)$ time by using the standard dynamic programming approach. In Section 2.4, we will show that by exploiting the underlying geometric properties of the DAG G, a K-link shortest path in G can be computed in $O(n2^{\sqrt{\log K \log\log n}})$ or $O(Kn)$ time.

2.3 Computing the Weights of the Edges in G

In this section, we show that for any l and j ($l < j$), the weight of the edge in G from u_l to u_j can be computed in $O(1)$ time after an $O(n)$ time preprocess. It suffices to consider the case when $x_j - x_{l+1} \leq \delta$. In this case, we have

$$\begin{aligned} V[l,j] &= \textstyle\sum_{i=l+1}^{j} P(x_i)(x_i - \mu[l,j])^2 \\ &= \textstyle\sum_{i=l+1}^{j} P(x_i)x_i^2 - 2\sum_{i=l+1}^{j} P(x_i) * x_i * \mu[l,j] + \sum_{i=l+1}^{j} P(x_i)(\mu[l,j])^2 \\ &= \textstyle\sum_{i=l+1}^{j} P(x_i)x_i^2 - \frac{(\sum_{i=l+1}^{j} P(x_i)x_i)^2}{\sum_{i=l+1}^{j} P(x_i)} \end{aligned}$$

Therefore, if we precompute all prefix sums of $\sum_{i=1}^{g} P(x_i)$, $\sum_{i=1}^{g} P(x_i)x_i$, and $\sum_{i=1}^{g} P(x_i)x_i^2$, $g = 1, 2, \ldots, n$, which can be easily done in $O(n)$ time, then we can compute any $V[l,j]$ in $O(1)$ time.

2.4 The Staircase-Monge Property of the Problem

This section shows that the discrete constrained 1-D K-means clustering problem satisfies the staircase-Monge property [2].

Lemma 1. *(1a) If $V[l,j] = +\infty$, then $V[l,j'] = +\infty$ for any $j' > j$ and $V[l',j] = +\infty$ for any $l' < l$.*
(1b) For any $0 < l+1 < j < n$, if the four entries $V[l,j], V[l+1,j], V[l,j+1]$, and $V[l+1,j+1]$ are all finite, then $V[l,j]+V[l+1,j+1] \leq V[l+1,j]+V[l,j+1]$.

Proof. (1a) If $V[l,j] = +\infty$, then $x_j - x_{l+1} > \delta$. Since the sequence of $x = (x_1, x_2, \ldots, x_n)$ is in ascending order, for any $j' > j$, we have $x_{j'} - x_{l+1} > x_j - x_{l+1} > \delta$. Hence $V[l,j'] = +\infty$. We can similarly argue that for any $l' < l$, $V[l',j] = +\infty$ holds.
(1b) Fix l and j. We can view $V[l+c, j+d]$'s and $\mu[l+c, j+d]$'s ($c = 0, 1$ and $d = 0, 1$) as multi-variable functions of $x_{l+1}, x_{l+2}, \ldots, x_{j+1}$ and $P(x_{l+1})$, $P(x_{l+2})$, ..., $P(x_{j+1})$. Let $W = V[i+1,j] + V[i,j+1] - V[i,j] - V[i+1,j+1]$. It is sufficient to show $W \geq 0$.