MATHEMATICAL ENGINEERING TECHNICAL REPORTS

Computational Geometric Approach to Submodular Function Minimization for Multiclass Queueing Systems

Toshinari Itoko and Satoru Iwata

METR 2005–29 October 2005

DEPARTMENT OF MATHEMATICAL INFORMATICS
GRADUATE SCHOOL OF INFORMATION SCIENCE AND TECHNOLOGY
THE UNIVERSITY OF TOKYO
BUNKYO-KU, TOKYO 113-8656, JAPAN

WWW page: http://www.i.u-tokyo.ac.jp/mi/mi-e.htm

Computational Geometric Approach to Submodular Function Minimization for Multiclass Queueing Systems

Toshinari Itoko *　　　Satoru Iwata †

Abstract

This paper presents an efficient algorithm for minimizing a certain class of submodular functions that arise in analysis of multiclass queueing systems. In particular, the algorithm can be used for testing whether a given multiclass M/M/1 system achieves a required average performance by an appropriate control policy. With the aid of the topological sweeping method for line arrangement, our algorithm runs in $O(n^2)$ time, where n is the cardinality of the ground set. This is much faster than direct applications of general submodular function minimization algorithms.

1 Introduction

Let V be a finite set of cardinality n. For a vector $\boldsymbol{x} := [x_i]_{i \in V}$ indexed by V and a subset $X \subseteq V$, we denote $\sum_{i \in X} x_i$ by $x(X)$. Let h be a nonnegative nondecreasing convex function. This paper deals with the problem of finding a subset $X \subseteq V$ that minimizes

$$f(X) := z(X) - y(X)\, h(x(X)) \qquad (X \subseteq V) \tag{1}$$

for given nonnegative vectors $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{z}$ indexed by V. Such a minimization problem arises in performance analysis of the most fundamental multiclass queueing system — multiclass M/M/1 (see Section 2).

This problem is a special case of submodular function minimization. A set function f is called *submodular* if it satisfies

$$f(X) + f(Y) \geq f(X \cap Y) + f(X \cup Y), \qquad \forall X, Y \subseteq V.$$

It can be shown that the function f in (1) is submodular (see Appendix). Recent results on submodular functions are expounded in Fujishige [9] and in McCormick [16].

A number of strongly polynomial algorithms have been devised for general submodular function minimization. The first one due to Grötschel, Lovász, and Schrijver [11, 12] is based on the ellipsoid method, which is not efficient in practice. Combinatorial strongly polynomial algorithms are devised independently by Schrijver [17] and by Iwata, Fleischer, and Fujishige [14]. However, these combinatorial algorithms are not yet very fast. Even an improved algorithm of Iwata [13], which is currently the fastest combinatorial strongly

*Tokyo Research Laboratory, IBM Japan, Yamato, Kanagawa 242-8502, Japan. itoko@jp.ibm.com

†Department of Mathematical Informatics, University of Tokyo, Tokyo 113-8656, Japan. Partially supported by the Okawa Foundation for Information and Telecommunications. iwata@mist.i.u-tokyo.ac.jp

polynomial algorithm, runs in $O((n^6\gamma + n^7)\log n)$ time, where γ is the time required for computing the function value of f. Thus, it is still desirable to have a fast algorithm for minimizing a specific class of submodular functions that naturally arise in applications.

Instead of applying an algorithm for general submodular function minimization, we take a completely different approach based on computational geometry. The first step is to interpret our problem in the three dimensional space as follows. Each subset $X \subseteq V$ corresponds to the point $(x(X), y(X), z(X)) \in \mathbb{R}^3_+$. The original problem is then equivalent to finding the minimum value of $\widehat{f}(x, y, z) := z - y\,h(x)$ among all such points (x, y, z) corresponding to the subsets of V.

The convex hull of these 2^n points forms a special polytope called a zonotope. It will be shown that the minimizer of $\widehat{f}$ is among the lower extreme points of Z, i.e., extreme points that are visible from below. The number of such lower extreme points are bounded by $O(n^2)$. Furthermore, exploiting the duality relation between the zonotope in the three dimensional space and the line arrangement in the plane, we are able to enumerate all the lower extreme points in $O(n^2)$ time with the aid of the topological sweeping method of Edelsbrunner and Guibas [5, 6]. Thus our algorithm finds a minimizer of f in $O(n^2)$ time and $O(n)$ space. This is substantially more efficient than direct applications of general submodular function minimization algorithms.

The outline of this paper is as follows. In Section 2, we provide a brief exposition on the multiclass M/M/1. Section 3 presents our minimization algorithm. In particular, Section 3.1 is devoted to the structure of the zonotope. In Section 3.2, we show that it suffices to compute the minimum among the lower extreme points. Then we discuss how to enumerate all the lower extreme points in Section 3.3.

2 Multiclass Queueing Systems

This section is devoted to a brief exposition on a connection between our minimization problem and the performance analysis of the fundamental multiclass queueing system called multiclass M/M/1. For comparison, we also give a brief description of the same type of problems for nonpreemptive case.

2.1 Preemptive M/M/1

Multiclass M/M/1 is a system which deals with various types of jobs whose arrival interval and service time follow exponential distributions. Each job of different classes wait in different queues and the server chooses the job to serve the next by a *control policy.* A queueing system allowing preemptive control policies is called *preemptive.* In the following, the set of classes is denoted by $V = \{1, 2, \ldots, n\}$.

In a multiclass M/M/1, when the average arrival rates and the average service rates of the job classes are given, the performance of the system depends only on the control policy. A region of performance-measuring vectors achieved by all control policies is called *achievable region* (see e.g. [3]). The performance of a multiclass M/M/1 is often measured by the average staying time vector $\boldsymbol{s} := [\,s_i\,]_{i\in V}$, where s_i is the average time in the system for class i jobs. For preemptive multiclass M/M/1, achievable region of $\boldsymbol{s}$ is known as follows.

Theorem 2.1 ([2]) Consider a preemptive multiclass M/M/1 whose average arrival rates are $\boldsymbol{\lambda}:= [\,\lambda_i\,]_{i\in V}$ and average service rates are $\boldsymbol{\mu}:= [\,\mu_i\,]_{i\in V}$. Let ρ_i be the utilization λ_i/μ_i of the server for class i jobs and assume $\sum_{i\in V}\rho_i < 1$ to ensure the existence of equilibrium. The achievable region of the average staying time vector $\boldsymbol{s} := [\,s_i\,]_{i\in V}$ is a polyhedron represented by 2^n inequalities:

$$\sum_{i\in X}\rho_i s_i \geq \frac{\displaystyle\sum_{i\in X}\frac{\rho_i}{\mu_i}}{1-\displaystyle\sum_{i\in X}\rho_i}, \quad \forall X\subseteq V. \tag{2}$$

Given a target average staying time vector $\check{\boldsymbol{s}}$, it is important for system designers to *check performance achievability*: whether $\check{\boldsymbol{s}}$ is in the achievable region ($\check{\boldsymbol{s}}$ is achieved by some control policy) or not. This problem was posed by Federgruen and Groenevelt [7]. They provided an efficient algorithm for the special case of identical service time distribution. This assumption is too restrictive in practice, as we usually classify the jobs by their properties including expected service time.

If we define $x_i := \rho_i$, $y_i := \dfrac{\rho_i}{\mu_i}$ and $h(x) := \dfrac{1}{1-x}$, then $y(X)\,h\big(x(X)\big)$ coincides with the right-hand side function of (2). Furthermore, if we define $z_i := \rho_i\check{s}_i$, then the problem of checking performance achievability of preemptive multiclass M/M/1 is reduced to our minimization problem. The target average staying time vector $\check{\boldsymbol{s}}$ is achievable if and only if the minimum value of f is equal to zero.

For preemptive multiclass M/M/1, there is an another representation of achievable region. Bertsimas, Paschalidis, and Tsitsiklis [1] and Kumar and Kumar [15] independently observed that the achievable region is the projection of a polyhedron in higher dimensional space. This makes it possible to check the achievability by solving linear programming problem, which however involves $O(n^2)$ variables and $O(n^2)$ inequalities.

2.2 Nonpreemptive M/G/1

For nonpreemptive M/M/1, which does not allow preemption, the performance achievability can be tested by a simpler method. The achievable region of the average waiting time in queue $\boldsymbol{q} := [\,q_i\,]_{i\in V}$ is a polyhedron represented by 2^n inequalities:

$$\sum_{i\in X}\rho_i\, q_i \geq \left(\sum_{i\in V}\frac{\rho_i}{\mu_i}\right)\frac{\displaystyle\sum_{i\in X}\rho_i}{1-\displaystyle\sum_{i\in X}\rho_i}, \qquad \forall X\subseteq V.$$

This is obtained as a special case of the fact shown in Gelenbe and Mitrani [10] that the achievable region of the nonpreemptive M/G/1, which admits general service time distributions, is characterized by

$$\sum_{i\in X}\rho_i\, q_i \geq \left(\frac{1}{2}\sum_{i\in V}\lambda_i M_i^2\right)\frac{\displaystyle\sum_{i\in X}\rho_i}{1-\displaystyle\sum_{i\in X}\rho_i}, \qquad \forall X\subseteq V,$$

where M_i^2 denotes the second order moment of the service time distribution for class i.

The problem of checking performance achievability is reduced to minimizing a submodular function b in the form of

$$b(X) := z(X) - h(x(X)) \qquad (X \subseteq V),$$

where $x_i := \rho_i$, $z_i := \rho_i q_i$, and $h(x) := \dfrac{c\,x}{1-x}$ with $c = \dfrac{1}{2}\sum_{i\in V}\lambda_i M_i^2$. This is much simpler than our problem. In fact, it can be solved by sorting job classes in the order of z_i/x_i. For $k = 0, 1, \ldots, n$, let Y_k denote the set of k jobs with smallest values of z_i/x_i. Then the minimizer of b is among the candidates Y_k for $k = 0, 1, \ldots, n$. See Federgruen and Groenevelt [8] for validity.

3 Geometric Approach

In this section, we present an algorithm for finding a minimum value of f defined in (1). The problem can be seen in the three-dimensional space as follows. A subset $X \subseteq V$ corresponds to a point $(x(X), y(X), z(X))$ in three-dimensional space. Let $\mathbb{R}_+$ denote the set of nonnegative reals. We also write $\boldsymbol{w}(X) := (x(X), y(X), z(X)) \in \mathbb{R}_+^3$ and $W := \{\boldsymbol{w}(X) \mid X \subseteq V\}$. Then our problem is equivalent to finding a point $(x, y, z) \in W$ that attains the minimum value of $\widehat{f}(x, y, z) = z - y\,h(x)$. An example of an contour surface of $\widehat{f}$ is shown in Figure 1.

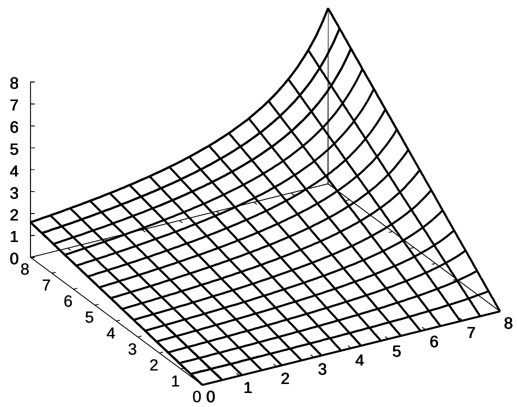

Figure 1: A contour surface of $\widehat{f}$ in the case of $h(x) = 2/(10 - x)$.

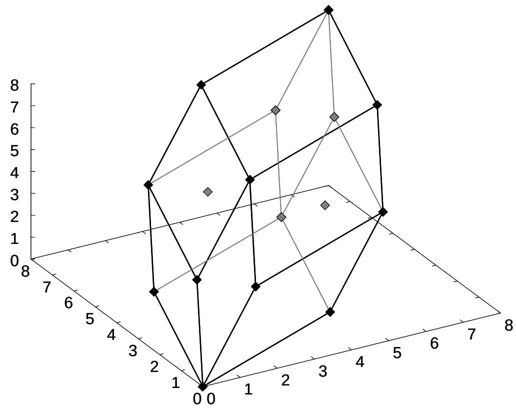

Figure 2: A zonotope generated by $(4, 1, 1)$, $(2, 1, 3)$, $(1, 2, 3)$, $(1, 4, 1)$.

The convex hull of W forms a special polytope called *zonotope*, which is defined by the bounded linear combination of vectors (for example, see Figure 2), namely

$$\mathrm{conv}(W) = \left\{ \sum_{i\in V} \eta_i\, \boldsymbol{w}_i \;\middle|\; 0 \le \eta_i \le 1 \ (\forall i \in V) \right\},$$

where $\boldsymbol{w}_i := \boldsymbol{w}(\{i\})$ $(\forall i \in V)$. We denote the convex hull of W by Z.

A point (x, y, z) in Z is called a lower point if $(x, y, z') \notin Z$ for any $z' < z$. If in addition (x, y, z) is a critical point of Z, it is called a lower extreme point. The number of lower extreme points of Z is known to be $O(n^2)$, which is clarified in Section 3.3.

Our algorithm enumerates all the lower extreme points of Z, and then it identifies a lower extreme point that attains the minimum value of $\widehat{f}$. It will be shown in Section 3.2 that the minimum value among these lower extreme points is in fact the minimum value of $\widehat{f}$ among all the points in W. How to enumerate the lower extreme points will be described in Section 3.3. The total running time of this algorithm is $O(n^2)$.

3.1 Lower Extreme Points

Every lower point of Z is described as a maximizer of a certain linear objective function whose coefficient of z is negative. For any $\alpha, \beta \in \mathbb{R}$, we denote by $F(\alpha, \beta)$ the set of maximizers for $(\alpha, \beta, -1)$ direction, namely

$$F(\alpha, \beta) := \mathrm{Argmax}\{\alpha\, x + \beta\, y - z \mid (x, y, z) \in Z\}.$$

For a fixed (α, β), elements in V are classified by the sign of $\alpha\, x_i + \beta\, y_i - z_i$, namely

$$\begin{array}{rcl} S^+(\alpha, \beta) & := & \{\, i \in V \mid \alpha\, x_i + \beta\, y_i - z_i > 0\,\}, \\ S^\circ(\alpha, \beta) & := & \{\, i \in V \mid \alpha\, x_i + \beta\, y_i - z_i = 0\,\}, \\ S^-(\alpha, \beta) & := & \{\, i \in V \mid \alpha\, x_i + \beta\, y_i - z_i < 0\,\}. \end{array} \tag{3}$$

Then $F(\alpha, \beta)$ is characterized by

$$F(\alpha, \beta) = \left\{ \boldsymbol{w}(S^+(\alpha, \beta)) + \sum_{i \in S^\circ(\alpha, \beta)} \eta_i\, \boldsymbol{w}_i \,\middle|\, \forall i \in S^\circ(\alpha, \beta),\ 0 \le \eta_i \le 1 \right\}. \tag{4}$$

This implies the following lemma that characterizes the lower extreme points of Z.

Lemma 3.1 A vector $\boldsymbol{v}$ is a lower extreme point of Z if and only if $\boldsymbol{v} = \boldsymbol{w}(S^+(\alpha, \beta))$ for some (α, β).

Proof. Since $\boldsymbol{w}_i \ge 0$, it follows from (4) that $\boldsymbol{w}(S^+(\alpha, \beta))$ is an extreme point of $F(\alpha, \beta)$. Hence $\boldsymbol{w}(S^+(\alpha, \beta))$ is an lower extreme point of Z. Conversely, suppose $\boldsymbol{v}$ is an lower extreme point of Z. There exists a pair (α, β) such that $\boldsymbol{v}$ is the unique maximizer of $\alpha x + \beta y - z$ in Z. Note that $\boldsymbol{v} = \boldsymbol{w}(X)$ for some $X \subseteq V$. Then we have $X = S^+(\alpha, \beta) \cup Y$ for some $Y \subseteq S^\circ(\alpha, \beta)$. Furthermore, since $\boldsymbol{v}$ is the unique maximizer, $\boldsymbol{w}_i = 0$ holds for any $i \in Y$, which implies $\boldsymbol{w}(X) = \boldsymbol{w}(S^+(\alpha, \beta))$. □

We henceforth denote $\{S^+(\alpha, \beta) \mid \alpha, \beta \in \mathbb{R}\}$ by $\mathcal{L}$. Then Lemma 3.1 asserts that the set of lower extreme points are given by $\{\boldsymbol{w}(X) \mid X \in \mathcal{L}\}$. The following two lemmas concerning lower points of Z will be used in the proof of the validity of our algorithm in Section 3.2.

Lemma 3.2 Any lower point $\boldsymbol{v}$ that is on an edge of Z is a convex combination of two lower extreme points $\boldsymbol{w}(X_1)$ and $\boldsymbol{w}(X_2)$ with $X_1 \subseteq X_2$.

Proof. There exists a pair (α, β) such that $F(\alpha, \beta)$ is the edge that contains $\boldsymbol{v}$. Then it follows from (4) that $F(\alpha, \beta)$ is a line segment between $\boldsymbol{w}(X_1)$ and $\boldsymbol{w}(X_2)$, where $X_1 = S^+(\alpha, \beta)$ and $X_2 = S^+(\alpha, \beta) \cup S^\circ(\alpha, \beta)$. □

Lemma 3.3 Any lower point $\boldsymbol{v}$ of Z is a convex combination of some at most three lower extreme points $\boldsymbol{w}(X_0)$, $\boldsymbol{w}(X_1)$, and $\boldsymbol{w}(X_2)$ with $X_0 \subseteq X_1 \subseteq X_2$.

Proof. There exists a pair (α, β) such that $F(\alpha, \beta)$ is the minimal face that contains $\boldsymbol{v}$. Then $\boldsymbol{w}(X_0)$ with $X_0 = S^+(\alpha, \beta)$ is an extreme point of $F(\alpha, \beta)$. Let $\boldsymbol{u}$ be the intersection between the half line from $\boldsymbol{w}(X_0)$ through $\boldsymbol{v}$ and the boundary of $F(\alpha, \beta)$. Note that $\boldsymbol{v}$ is a convex combination of $\boldsymbol{w}(X_0)$ and $\boldsymbol{u}$. Since $\boldsymbol{u}$ is on an edge of Z, Lemma 3.2 implies that $\boldsymbol{u}$ is a convex combination of lower extreme points $\boldsymbol{w}(X_1)$ and $\boldsymbol{w}(X_2)$ with $X_1 \subseteq X_2$. Furthermore, since $\boldsymbol{w}(X_1)$ and $\boldsymbol{w}(X_2)$ are extreme points of $F(\alpha, \beta)$, we have $X_0 \subseteq X_1, X_2$. Therefore, $\boldsymbol{v}$ is a convex combination of $\boldsymbol{w}(X_0)$, $\boldsymbol{w}(X_1)$, and $\boldsymbol{w}(X_2)$ with $X_0 \subseteq X_1 \subseteq X_2$. □

3.2 Finding the Minimum Value

The following theorem shows that it suffices to examine the lower extreme points of Z on behalf of the points in W. This leads us to an efficient algorithm for finding the minimum value of f, provided that an enumeration algorithm for the lower extreme points is available.

Theorem 3.4 The minimum value of f is attained by a member of $\mathcal{L}$, i.e.,

$$\min\{f(X) \mid X \subseteq V\} = \min\{f(X) \mid X \in \mathcal{L}\}.$$

Proof. Let $\bar{\boldsymbol{v}} = (\bar{x}, \bar{y}, \bar{z})$ be a lower point of Z such that $\bar{x} = x(Y)$ and $\bar{y} = y(Y)$ for $Y \subseteq V$. By Lemma 3.3, there exist three lower extreme points $\boldsymbol{w}(X_0)$, $\boldsymbol{w}(X_1)$, and $\boldsymbol{w}(X_2)$ of Z with $X_0 \subseteq X_1 \subseteq X_2$ such that

$$\bar{\boldsymbol{v}} = \sigma_0 \boldsymbol{w}(X_0) + \sigma_1 \boldsymbol{w}(X_1) + \sigma_2 \boldsymbol{w}(X_2)$$

for some $\sigma_0, \sigma_1, \sigma_2 \geq 0$ with $\sigma_0 + \sigma_1 + \sigma_2 = 1$. We denote $\boldsymbol{w}(X_j)$ by (x_j, y_j, z_j) for $j = 0, 1, 2$. Then we have

$$\begin{aligned}
\bar{y}\, h(\bar{x}) &= (\sigma_0 y_0 + \sigma_1 y_1 + \sigma_2 y_2)\, h(\sigma_0 x_0 + \sigma_1 x_1 + \sigma_2 x_2) \\
&\leq (\sigma_0 y_0 + \sigma_1 y_1 + \sigma_2 y_2)(\sigma_0 h(x_0) + \sigma_1 h(x_1) + \sigma_2 h(x_2)) \\
&= \sigma_0 y_0 h(x_0) + \sigma_1 y_1 h(x_1) + \sigma_2 y_2 h(x_2) - \sigma_0 \sigma_1 (y_1 - y_0)(h(x_1) - h(x_0)) \\
&\quad - \sigma_1 \sigma_2 (y_2 - y_1)(h(x_2) - h(x_1)) - \sigma_0 \sigma_2 (y_2 - y_0)(h(x_2) - h(x_0)) \\
&\leq \sigma_0 y_0 h(x_0) + \sigma_1 y_1 h(x_1) + \sigma_2 y_2 h(x_2),
\end{aligned}$$

where the first inequality follows from the convexity of h and the second one from the monotonicity. Since $z(Y) \geq \bar{z} = \sigma_0 z_0 + \sigma_1 z_1 + \sigma_2 z_2$, we obtain

$$\begin{aligned}
f(Y) &= z(Y) - y(Y)\, h(x(Y)) \\
&\geq \bar{z} - \bar{y}\, h(\bar{x}) \\
&\geq \sigma_0 (z_0 - y_0 h(x_0)) + \sigma_1 (z_1 - y_1 h(x_1)) + \sigma_2 (z_2 - y_2 h(x_2)) \\
&= \sigma_0 f(X_0) + \sigma_1 f(X_1) + \sigma_2 f(X_2).
\end{aligned}$$

Therefore, if $f(Y)$ attains the minimum value, then X_0, X_1, and X_2 must attain the minimum value as well. Thus the minimum value of f is attained by a member of $\mathcal{L}$. □

3.3 Duality between Zonotope and Hyperplane Arrangement

In this section, we discuss how to enumerate all the lower extreme points of Z. A one-to-one correspondence has been established between zonotopes in the d dimensional space and hyperplane arrangements in the $d-1$ dimensional space (see e.g. [4, 18]). We exploit the three dimensional case of this duality principle.

To visualize the transition of $S^+(\alpha, \beta)$ in $\mathcal{L}$ with respect to α and β, we consider the arrangement of n lines $l_i : \alpha\, x_i + \beta\, y_i - z_i = 0$, for $i \in V$ in the (α, β)-plane. Then it follows from Lemma 3.1 that the lower extreme points of Z corresponds to the cells in this line arrangement. Note that the number of cells in the line arrangement is $O(n^2)$, and so is the number of lower extreme points of Z. Further correspondence between the lower faces of Z and the components of the line arrangement are summarized in Table 1 (see also Figure 3).

Table 1: Correspondence between lower faces of Z and components of line arrangement.

Lower faces of Z	Components of line arrangement
extreme point	cell
edge	line segment
facet	intersection point

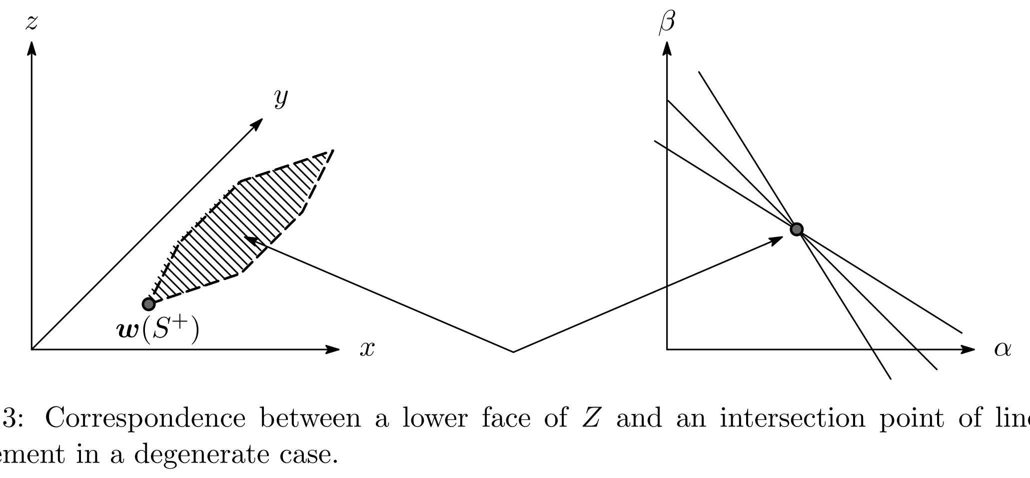

Figure 3: Correspondence between a lower face of Z and an intersection point of line arrangement in a degenerate case.

Based on this duality, it suffices to enumerate all the cells of the line arrangement. Sweeping algorithms for line arrangement keep a chain on V that corresponds to n cells and enumerate all the cells one by one with updating the chain. Our algorithm maintain not only the chain but also the vectors $\boldsymbol{w}(S^+)$ for all S^+ in the chain to compute the value of $f(S^+)$ in a constant time on average. This is achieved by minor modifications of existing algorithms at no additional expense of running time bound and space complexity.

In theory, the topological sweeping method [5, 6] is the most efficient algorithm, which runs in $O(n^2)$ time and $O(n)$ space. Thus we obtain an algorithm to solve the minimization problem in $O(n^2)$ time and $O(n)$ space.

4 Conclusion

We have presented an efficient algorithm for minimizing a class of submodular function that arises in queueing analysis: checking performance achievability of preemptive multiclass M/M/1. Employing the topological sweeping method for line arrangement, our algorithm runs in $O(n^2)$ time, which is much faster than previously known methods. Even if our algorithm discerns that the target performance is achievable, it does not yield a concrete control policy that achieves such a performance. Developing a fast algorithm for finding a control policy in the achievable case is left for future research.

Acknowledgments

We thank Takeshi Tokuyama of Tohoku University for helpful information on topological sweeping method. We also thank Takayuki Osogami of Tokyo Research Laboratory, IBM Japan, for helpful information on performance analysis of priority queues.

References

[1] D. Bertsimas, I. Ch. Paschalidis, and J. N. Tsitsiklis. Optimization of multiclass queueing networks: Polyhedral and nonlinear characterizations of achievable performance. *The Annals of Applied Probability*, 4:43–75, 1994.

[2] E. G. Coffman, Jr. and I. Mitrani. A characterization of waiting time performance realizable by single-server queues. *Operations Research*, 28:810–821, 1980.

[3] M. Dacre, K. Glazebrook, and J. Niño-Mora. The achievable region approach to the optimal control of stochastic systems. *Journal of the Royal Statistical Society, B*, 61:747–791, 1999.

[4] H. Edelsbrunner. *Algorithms in Combinatorial Geometry.* Springer-Verlag, 1987.

[5] H. Edelsbrunner and L. J. Guibas. Topologically sweeping an arrangement. *Journal of Computer and Systems Sciences*, 38:165–194, 1989.

[6] H. Edelsbrunner and L. J. Guibas. Topologically sweeping an arrangement — a correction. *Journal of Computer and Systems Sciences*, 42:249–251, 1991.

[7] A. Federgruen and H. Groenevelt. Characterization and optimization of achievable performance in general queueing systems. *Operations Research*, 36:733–741, 1988.

[8] A. Federgruen and H. Groenevelt. M/G/c queueing systems with multiple customer classes: Characterization and control of achievable performance under nonpreemptive priority rules. *Management Science*, 34:1121–1138, 1988.

[9] S. Fujishige. *Submodular Function and Optimization.* North-Holland, 2005.

[10] E. Gelenbe and I. Mitrani. *Analysis and Synthesis of Computer Systems.* Academic Press, 1980.

[11] M. Grötschel, L. Lovász, and A. Schrijver. The ellipsoid method and its consequences in combinatorial optimization. *Combinatorica*, 1:169–197, 1981.

[12] M. Grötschel, L. Lovász, and A. Schrijver. *Geometric Algorithms and Combinatorial Optimization.* Springer-Verlag, 1988.

[13] S. Iwata. A faster scaling algorithm for minimizing submodular functions. *SIAM Journal on Computing*, 32:833–840, 2003.

[14] S. Iwata, L. Fleischer, and S. Fujishige. A combinatorial strongly polynomial algorithm for minimizing submodular functions. *Journal of the ACM*, 48:761–777, 2001.

[15] S. Kumar and P. R. Kumar. Performance bounds for queueing networks and scheduling policies. *IEEE Transactions on Automatic Control*, 39:1600–1611, 1994.

[16] S. T. McCormick. Submodular function minimization. In K. Aardal, G. Nemhauser, and R. Weismantel, editors, *Handbook on Discrete Optimization.* Elsevier, to appear.

[17] A. Schrijver. A combinatorial algorithm minimizing submodular functions in strongly polynomial time. *Journal of Combinatorial Theory, B*, 80:346–355, 2000.

[18] G. M. Ziegler. *Lectures on Polytopes.* Springer-Verlag, 1995.

Appendix

This Appendix is devoted to show that the function f is a submodular function. For this purpose, it suffices to show that the function g defined by $g(X) := y(X)\, h(x(X)$ is supermodular, i.e.,

$$g(X) + g(Y) \leq g(X \cap Y) + g(X \cup Y), \qquad \forall X, Y \subseteq V.$$

Since h is convex, for any $X \subseteq V$ and $i, j \in V$, we have

$$\begin{aligned} h\big(x(X \cup \{i\})\big) &\leq \frac{x_j}{x_i + x_j}\, h\big(x(X)\big) + \frac{x_i}{x_i + x_j}\, h\big(x(X \cup \{i, j\})\big), \\ h\big(x(X \cup \{j\})\big) &\leq \frac{x_i}{x_i + x_j}\, h\big(x(X)\big) + \frac{x_j}{x_i + x_j}\, h\big(x(X \cup \{i, j\})\big). \end{aligned}$$

By adding these two inequalities, we obtain

$$h\big(x(X \cup \{i\})\big) + h\big(x(X \cup \{j\})\big) \leq h\big(x(X)\big) + h\big(x(X \cup \{i, j\})\big),$$

which implies that $h(x(\cdot))$ is a supermodular function. Because of this supermodularity, the nonnegativity of x and y, and the monotonicity of h, we have

$$\begin{aligned} & g(X \cup Y) + g(X \cap Y) - g(X) - g(Y) \\ &= h(x(X \cup Y))\, y(X \cup Y) + h(x(X \cap Y))\, y(X \cap Y) - h(X)\, y(X) - h(Y)\, y(Y) \\ &= \big(h(x(X \cup Y)) + h(x(X \cap Y)) - h(x(X)) - h(x(Y))\big)\, y(X \cap Y) \\ &\quad + \big(h(x(X \cup Y)) - h(X)\big)\, y(X \setminus Y) + \big(h(x(X \cup Y)) - h(Y)\big)\, y(Y \setminus X) \\ &\geq 0 \end{aligned}$$

for any $X, Y \subseteq V$. Thus g is shown to be supermodular. In addition, it is easy to see that $g(X) \geq 0$ for any $X \subseteq V$ and $g(\emptyset) = 0$ hold.