Multimodularity, Convexity and Optimization Properties*

Eitan ALTMAN[†] Bruno GAUJAL[‡] Arie HORDIJK[§]

May 26, 1999

Abstract

We investigate in this paper the properties of multimodular functions. In doing so we give elementary proofs for properties already established by Hajek, and we generalize some of his results. In particular, we extend the relation between convexity and multimodularity to some convex subsets of $\mathbb{Z}^m$. We also obtain general optimization results for average costs related to a sequence of multimodular functions rather than to a single function. Under this general context, we show that the expected average cost problem is optimized by using balanced sequences. We finally illustrate the usefulness of this theory in admission control into a D/D/1 queue with fixed batch arrivals, with no state information. We show that the balanced policy minimizes the average queue length for the case of an infinite queue, but not for the case of a finite queue. When further adding a constraint on the losses, it is shown that a balanced policy is also optimal for the finite queue case.

Keywords Multimodular functions, convexity, balanced sequences, admission control into a queue.

1 Introduction

The multimodularity property of functions was much investigated in the context of queuing systems. There are several cases in that field in which this property was exploited in

*The work was supported by the Van Gogh project N. 98001 and INRIA action coopérative Maddes

[†]INRIA, 2004 Route des Lucioles, BP 93, 06902 Sophia Antipolis Cedex, France. E-mail: altman@sophia.inria.fr. URL:http://www-sop.inria.fr/mistral/personnel/Eitan.Altman/me.html

[‡]LORIA, 615, rue du jardin botanique, BP 101, 54602 Villers lès Nancy Cedex, France. E-mail: Bruno.Gaujal@loria.fr. URL:http://www.loria.fr/~gaujal

[§]Dept. of Mathematics and Computer Science, Leiden University, P.O.Box 9512, 2300RA Leiden, The Netherlands. E-mail: hordijk@wi.leidenuniv.nl. The research of Arie Hordijk was initiated while he was on sabbatical leave at INRIA, Sophia-Antipolis; partially supported by the Ministère Français de l'Éducation Nationale et de l'Enseignement Supérieur et de la Recherche.

stochastic control. Optimal admission control under no queue information was studied by Hajek [6]. The precise problem was to admit customers to a single queue, under the constraint that the long run fraction of customers admitted be at least p. The optimality of a policy based on a balanced sequence of admission actions was obtained in [6] for the number of customers in a one-server queue with exponential service and a renewal arrival process.

Another application of multimodular functions is in the control of queues with full state information. Weber and Stidham [13] (and later Glasserman and Yao [5]) obtained monotone properties of the optimal control policies as a function of the state, in a problem of control of service rates in a system of m queues in tandem. The methodology was strongly based on the multimodularity properties of the immediate costs and the cost-to-go functions.

The first purpose of this paper is to study the properties of multimodular functions, as a generic tool for further investigating the control of queuing systems. We provide elementary proofs for properties already established by Hajek, who showed that the lower convex envelope of a multimodular function f is the piecewise linear interpolation on a specific triangulation of the space by simpleces called atoms. In this paper, we show directly that this linear interpolation is convex if and only if f is multimodular. This allows us to restrict the study of multimodular functions to convex subsets of $\mathbb{Z}^m$ which are convex unions of atoms (or faces of atoms). Additional interesting properties of multimodular functions are also presented.

In the second part of the paper, we develop basic optimization tools for average costs. We establish lower bounds for average costs using Abel-type asymptotic techniques. We also show that the lower bounds are achieved by balanced sequences. Such costs depend on a sequence of multimodular functions, rather than on a single multimodular function. This is a nice feature of our approach since the optimization results can be applied directly to average costs as long as the assumptions used in Theorems 3.1 or 3.2 are satisfied. This is not the case when a single function is used, as in [6], where for any specific application, additional analysis has to be done before one may apply the general minimization results to average costs problems.

We finally illustrate the usefulness of this theory in admission control into a queue; we cite some results for the G/G/1 queue, and provide a detailed analysis of the D/D/1 queue with fixed batch arrivals, with no state information. We show, for the latter, that the policy which is defined through a balanced sequence minimizes the average queue length for the case of an infinite queue, but not for the case of a finite buffer. However, when further restricting to those policies for which no losses occur, we obtain again the optimality of balanced policies. To conclude that example, we study also the case where it is possible to admit a part of an arriving batch.

In follow-up papers [1, 2, 3], we shall make use of all the theoretical results of this paper in order to study more general admission and service control problems in dynamic systems that can be described using the max-plus algebra, with general stationary inter-arrival and service times.

2 Properties of multimodular functions

We present in this section a short overview of Hajek's theory of multimodular functions. Some additional results are also established. We begin by presenting the definition of multimodularity, and some general properties (Subsection 2.1) which have an interest by their own and that will be used in subsequent work. We then present in Subsection 2.2 the relation between multimodularity and convexity. The properties presented in Subsection 2.2 are those needed in the following sections and subsequent papers on optimization and control.

Let $e_i \in \mathbb{N}^m$, $i = 1, ..., m$ denote the vector having all entries zero except for a 1 in its ith entry. Define $d_i = e_{i-1} - e_i$, $i = 1, ..., m$ (for an integer i taking values between 0 and m, we understand throughout $i - 1 = m$ for $i = 0$).

Let $\mathcal{F} = \{-e_1, d_2, ..., d_m, e_m\}$. Define $\mathcal{G} = \{e_i, -e_i, d_i, -d_i,\ i = 0, 1, ..., m\}$.

Definition 2.1 (Hajek). *A function f on $\mathbb{Z}^m$ is multimodular with respect to $\mathcal{F}$ if for all $x \in \mathbb{Z}^m$, $v, w \in \mathcal{F}$, $v \neq w$,*

$$f(x+v) + f(x+w) \geqslant f(x) + f(x+v+w). \tag{1}$$

Unless otherwise stated, we shall say that f is multimodular if it is multimodular with respect to $\mathcal{F}$.

2.1 General properties

For a function g defined on $\mathbb{Z}^m$, define

$$\Delta_i g(x) = \Delta_{e_i} g(x) = g(x+e_i) - g(x) \qquad \textit{and} \qquad \Delta_{d_i} g = \Delta_{i-1} g - \Delta_i g.$$

We further define $\Delta_{-e_i} g = g(x - e_i) - g(x)$. Note that $\Delta_{d_i} g(x) = g(x + e_i + d_i) - g(x + e_i)$.

It is easy to check that

Lemma 2.1. *$\Delta_v g$ is a linear function for any $v \in \mathcal{G}$. For all $v, w \in \mathcal{G}$, $\Delta_v \Delta_w g = \Delta_w \Delta_v g$.*

Lemma 2.2. *(a) f is multimodular if and only if*

$$\Delta_v \Delta_w f \leqslant 0 \tag{2}$$

for all $v, w \in \mathcal{F}$, $w \neq v$.

(b) If f is multimodular then

- *(b.i) For any $w \in \mathcal{F}$ with $w \neq e_1$.*

$$\Delta_{e_1} \Delta_w f \geqslant 0. \tag{3}$$

- *(b.ii) For all i, j,*

$$\Delta_i \Delta_j f \geqslant 0. \tag{4}$$

- *(b.iii) For all i, j,*

$$\Delta_j \Delta_j f \geqslant \Delta_i \Delta_j f. \tag{5}$$

- *(b.iv) $\Delta_{e_i} \Delta_{d_i} f \geqslant 0$.*

- *(b.v) $\Delta_{e_i} \Delta_{d_j} f \leqslant 0$, $j < i$ and $\Delta_{e_i} \Delta_{d_j} f \geqslant 0$, $j > i$.*

- *(b.vi) $\Delta_{d_1} \Delta_{d_i} f \leqslant 0$, $i \neq 1$.*

- *(b.vii) $\Delta_{d_i} \Delta_{d_i} f \geqslant 0$.*

(c) Consider the 2-dimensional case: $\mathcal{F} = \{-e_1, d_2, e_2\}$. Assume that (4) and (5) hold. Then f is multimodular.

The proof of the lemma is technical and tedious. It is given in Appendix A.

Note that Equations (b.ii) and (b.iii) can be seen as a discrete counterpart of the characterization of convexity using second derivatives in the continuous domain. Equations (b.iv), (b.v), (b.vi) and (b.vii) are useful when dealing with functions of multimodular functions, like projections. Indeed, checking multimodularity of projections or restrictions of multimodular functions is easier using this approach.

For example, given a point X in $\mathbb{Z}^m$, a set I of indices $i_1 \leqslant \cdots \leqslant i_k$ with $k \leqslant n$ and the function $p_{X,I} : \mathbb{Z}^k \to \mathbb{Z}^m$ defined by $p_{X,I}(Y) = Z$, with $Z_i = X_i$ if $i \notin I$ and $Z_{i_j} = Y_j$ otherwise, we have the following property.

Lemma 2.3. *If $f : \mathbb{Z}^m \to \mathbb{R}$ is multimodular then for all I and X, the function $g : \mathbb{Z}^k \to \mathbb{R}$ defined by $g(Y) = f(p_{X,I}(Y))$ is multimodular.*

Proof. The proof holds by checking that relations (2) holds for g. Using the previous notations, we have to check

$$\begin{aligned} \Delta_{-e_1} \Delta_{d_j} g(Y) &= \Delta_{-e_{i_1}} \Delta_{\sum_{t=i_{j-1}+1}^{i_j} d_t} f(Z) \\ &= \sum_{t=i_{j-1}+1}^{i_j} \Delta_{-e_{i_1}} \Delta_{d_t} f(Z) \\ &\leqslant 0, \end{aligned}$$

by Equation (b.v).

The other cases, $\Delta_{d_i} \Delta_{d_j} g(Y)$ and $\Delta_{d_i} \Delta_{e_k} g(Y)$ are checked similarly. ∎

2.2 Multimodularity and convexity

In the first part of this section, we present some details on the construction of a simplicial decomposition of $\mathbb{R}^m$ adapted to the base of multimodularity. This construction was given by Hajek. Theorem 4.3 in [6] proves that the lower convex envelope $\widetilde{f}$ of a multimodular function f is the piecewise linear interpolation of f on this triangulation. In this paper we define $\widetilde{f}$ directly as the piecewise linear interpolation of f. The second part of the section is then devoted to Theorem 2.1, that shows that f is multimodular if and only if $\widetilde{f}$ is convex, and some of its consequences which are useful to derive optimality results (see Section 5).

We first introduce the notion of *atoms*, which was used by Hajek. In the space $\mathbb{R}^m$, the convex hull of $m+1$ linearly independent points in $\mathbb{Z}^m$ forms a simplex. A simplex defined on the set of points $\{x^0, ..., x^m\}$ of $\mathbb{Z}^m$ is called an *atom* (defined in [6] §3) if and only if for some ordering of the set and for some permutation $(i_0, \cdots, i_m)$ of $(0, 1, ..., m)$,

$$\begin{array}{rcr} x^1 & = & x^0 + g_{i_1} \\ x^2 & = & x^1 + g_{i_2} \\ & \cdot & \\ & \cdot & \\ & \cdot & \\ x^m & = & x^{m-1} + g_{i_m} \\ x^0 & = & x^m + g_{i_0} \end{array} \tag{6}$$

where $g_{i_0}, ..., g_{i_m}$ are the elements of $\mathcal{F}$.

Next we present a characterization of an atom (see [6]), which is essential for the optimization result that we obtain in the following sections. Denote by $\lfloor x \rfloor$ the largest integer smaller than or equal to x. Then the following trivially holds

$$\int_0^1 \lfloor x + \theta \rfloor d\theta = x. \tag{7}$$

Given $z \in \mathbb{R}^m$, $\theta \in \mathbb{R}$, define the vector $u^z(\theta)$ in $\mathbb{Z}^m$:

$$u_i^z(\theta) = \lfloor \theta + z_1 + ... + z_i \rfloor - \lfloor \theta + z_1 + ... + z_{i-1} \rfloor,$$

$i = 1, ..., m$. Then by (7),

$$\int_0^1 u^z(\theta) d\theta = z. \tag{8}$$

$u^z(\theta)$ is periodic in θ with period 1, and piecewise constant with at most $m+1$ jumps per period. Thus, the set $\{u^z(\theta) : 0 \leqslant \theta \leqslant 1\}$ contains at most $m+1$ vectors, all integer valued. The next Lemma follows from [6]:

Lemma 2.4. *A point z is contained in an atom, say $S(z)$, if and only if the extreme points of $S(z)$ contain $\{u^z(\theta) : 0 \leqslant \theta \leqslant 1\}$. A point z is in the interior of atom $S(z)$ if and only if the extreme points of $S(z)$ equal $\{u^z(\theta) : 0 \leqslant \theta \leqslant 1\}$.*

Note that z may be a point in the intersection of two or more atoms. But each point $z \in \mathbb{R}^m$ is contained in some atom, say $S(z)$, and it can thus be expressed as the convex combination given by (8) of the extreme points of $S(z)$.

For any function f on $\mathbb{Z}^m$, we define the corresponding function $\widetilde{f}$ on $\mathbb{R}^m$ as follows. It agrees with f on $\mathbb{Z}^m$, and its value on an arbitrary point in $z \in \mathbb{R}^m$ is obtained as the corresponding linear interpolation of the values of f on the extreme points of the atom $S(z)$. Note that $\widetilde{f}$ is uniquely defined, if z belongs to atoms S_1 and S_2 then the points $u^z(\theta)$ belong to the extreme points of $S_1 \cap S_2$.

The following theorem establishes the equivalence between multimodularity of the discrete function and convexity of its continuous extension. In [6], the "only if" part was proved (this is the hard part), while the "if" part was omitted. In [4], the equivalence is established in the more general context of general multimodular triangulations. However, in both cases, the proofs are rather involved and use separating hyperplane techniques. The proof presented here is elementary and only uses a discrete counterpart of the second derivative argument. Furthermore, this proof technique enables us to prove this equivalence on special convex subsets of $\mathbb{Z}^m$, as shown in Corollary 2.1.

Theorem 2.1. *f is multimodular if and only if $\widetilde{f}$ is convex.*

Proof. <u>"only if"</u>: The function $\widetilde{f}$ is continuous by definition. Moreover, along any direction d, it has only a discrete number of isolated points where it is not differentiable. By using the characterization of convexity given in [10], we will check convexity at a point z by showing that for point z, and any direction d, the right derivative is greater than or equal to the left derivative. It obviously suffices to check at points that are on the boundary of an atom, since, by definition, $\widetilde{f}$ is linear in the interior of atoms. Hence, we first assume that the point z is on the interior of a face (of dimension $m-1$) which is common between two adjacent atoms. Without loss of generality, assume that the atoms (defined below by their extreme points) are

$$A = A(x_0, x_1, ..., x_m) \qquad \text{and} \qquad \overline{A} = A(x_0, x_1^*, ..., x_m).$$

where x_i satisfy (6) and

$$x_1^* = x_0 + g_{i_2}, \qquad x_2 = x_1^* + g_{i_1}.$$

<u>Case 1:</u> $g_{i_1} = -e_1 = (-1, 0, ..., 0)$, $g_{i_2} = e_m = (0, 0, ..., 1)$.

Decompose direction d in its projection ∂_2 in the common face between the two atoms and in the component ∂_1 along the direction $(x_1^* - x_1)$. In the direction ∂_2, the left and right derivatives are equal. In the direction ∂_1, the right derivative is a constant c, depending on the length of ∂_1, times $\widetilde{f}(x_1^*) - \widetilde{f}(z)$. The left derivative is $c(\widetilde{f}(z) - \widetilde{f}(x_1))$. Omitting the constant c, and using point $z = \frac{1}{2}(x_0 + x_2)$ hence $2\widetilde{f}(z) = f(x_0) + f(x_2)$, we get for the difference

$$\begin{aligned} &(\widetilde{f}(x_1^*) - \widetilde{f}(z)) - (\widetilde{f}(z) - \widetilde{f}(x_1)) \\ &= \ (f(x_1^*) - f(x_0)) - (f(x_2) - f(x_1)) \end{aligned} \tag{9}$$

The fact that (9) is nonnegative follows by applying (1) with $x = x_0$, and

$$x_1^* = x_0 + e_m$$
$$x_2 = x_0 - e_1 + e_m$$
$$x_1 = x_0 - e_1.$$

Case 2: $g_{i_1} = e_m$ and $g_{i_2} = -e_1$. It is handled as Case 1.
Case 3: $g_{i_1} = d_2 = (1, -1, 0, ..., 0)$, $g_{i_2} = -e_1$. We set $x = x_0$, and $x_1^* = x_0 - e_1, x_2 = x_0 - e_1 + d_2, x_1 = x_0 + d_2$. We decompose d along $-e_1$ on its projection ∂_2 in the common face, and in the projection ∂_1 along $(x_1^* - x_0)$. As in Case 1, it suffices to consider the direction ∂_1. The right derivative in this direction is $f(x_1^*) - f(x_0)$, and the left derivative is $f(x_2) - f(x_1)$ (both up to a multiplicative constant).

The difference between the right and left derivatives is indeed nonnegative: $f(x_1^*) - f(x_0) - (f(x_2) - f(x_1)) \geqslant 0$. This is obtained again by applying (1).

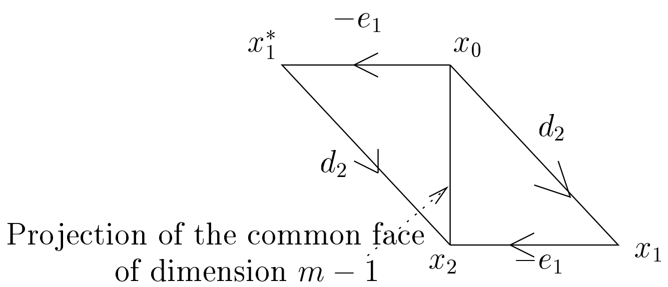

Figure 1: Checking convexity at a point on the common face, Case 3

Case 4: $g_{i_1} = d_2 = (1, -1, 0, ..., 0)$, $g_{i_2} = d_3$. In this case, we project d along $(x_1^* - x_1)$, and the analysis is as for Case 1.

All other cases, in which z is in the interior of a face (of dimension $m - 1$), common to two adjacent atoms, are similar to one of those considered above. It now remains to consider the case where the direction d in point z crosses from atom A to atom $\overline{A}$, and $A \cap \overline{A}$ is of dimension at most $m - 2$. In that case, we consider a cylinder C in direction d containing point z and with an arbitrarily small diameter. This cylinder intersects atoms A and $\overline{A}$ and is covered by atoms. We consider the projection P_C of C along direction d. This has dimension $m - 1$. The intersection (say of dimension k) of C with an atom is projected on the intersection of the projections and has dimension at most k. Therefore, P_C is almost everywhere (in Lesbesgue measure) covered with projections of dimension $m - 1$. We can find a line L that belongs to C and intersects A and $\overline{A}$, with direction d, and which projection is a point in P_C not belonging to intersections of dimension smaller than $m - 1$. Therefore, we can claim that L only intersects faces of atoms of dimension $m - 1$. The convexity in point z and direction d now follows from the convexity in points z_i corresponding to the intersections of line L with all the intermediate atoms between A and $\overline{A}$ and a continuity argument.

"if": Consider an arbitrary point x_0 and any two distinct elements g_i, g_j in $\mathcal{F}$. We have to show that

$$f(x_0) + f(x_2) - f(x_1) - f(x_1^*) \leqslant 0, \tag{10}$$

where $x_1 \stackrel{\triangle}{=} x_0 + g_i$, $x_1^* \stackrel{\triangle}{=} x_0 + g_j$, $x_2 \stackrel{\triangle}{=} x_1 + g_j = x_1^* + g_i$.

Define $z \stackrel{\triangle}{=} \frac{1}{2}(x_1 + x_1^*) = \frac{1}{2}(x_0 + x_2)$ and consider the line segment $x_1 \to z \to x_1^*$. The left derivative (l.d.) and right derivative (r.d.) in z are given by

$$l.d. = \widetilde{f}(\frac{1}{2}(x_1 + x_1^*)) - \widetilde{f}(x_1) = \frac{1}{2}f(x_0) + \frac{1}{2}f(x_2) - f(x_1),$$

$$r.d. = \widetilde{f}(x_1^*) - \widetilde{f}(\frac{1}{2}(x_0 + x_2)) = f(x_1^*) - \frac{1}{2}f(x_0) - \frac{1}{2}f(x_2).$$

Since $\widetilde{f}$ is convex, $r.d. - l.d.$ is non-negative, and hence (10) holds. ∎

We can restrict the notion of multimodularity to some convex sets of $\mathbb{Z}^m$. Indeed, let A be a convex set which is a union of a set of atoms (or of a faces of atoms). We restrict the definition of multimodularity to directions that only lead to points in A. More precisely, we say that f is multimodular in A if the following holds.

If $x_0, x_0 + g_i, x_0 + g_j, x_0 + g_i + g_j$ are all elements of A then

$$f(x_0) + f(x_0 + g_i + g_j) - f(x_0 + g_i) - f(x_0 + g_j) \leqslant 0.$$

Corollary 2.1. *The function f is multimodular in A if and only if $\widetilde{f}$ is convex on A.*

Proof. It should be clear from the proof of Theorem 2.1 that the equivalence of the multimodularity of f and the convexity of $\widetilde{f}$ still holds if we restrict the function f to A. ∎

The restriction to $A = \mathbb{N}^{m+1}$ turns out to be essential for the application of Theorem 2.1 presented in Section 5.1. Indeed, as mentioned in Remark 5.1, the function considered is only defined on non-negative coordinates. Also, in Section 5.2, an even smaller set A is used in order to consider a case of constrained optimization (see Lemma 5.1).

A second corollary of Theorem 2.1 concerns the minimization of multimodular functions. For a function defined on A, we call x a local minimum on A if $f(x) \leqslant f(x+/-e_i)$ for all i such that $x+/-e_i$ is in A.

Corollary 2.2. *Let the function f be multimodular in A. Then a local minimum is a global minimum on A.*

Proof. If f is multimodular in A, then $\widetilde{f}$ is convex in A, and is linear on the (faces of) atoms forming A. The graph of $\widetilde{f}$ (*i.e.* $\{x : \exists y \text{ s.t. } x \geqslant \widetilde{f}(y)\}$) is a convex polytope. Therefore, all the local minima are global and are extreme points of atoms. ∎

Next we consider the integer convexity properties of a function f. A function f is said to be integer convex if the following holds. For vectors x and d in $\mathbb{Z}^m$, we have

$$f(x + d) - f(x) \geqslant f(x) - f(x - d).$$

Theorem 2.2. *Let f be multimodular. Then it is integer convex.*

Proof. Define $\delta_d^+(x)$:= the right derivative of $\widetilde{f}$ at x in the direction d and $\delta_d^-(x)$:= the left derivative of $\widetilde{f}$ at x in the direction d. Since $\widetilde{f}$ is convex (Theorem 2.1) then

$$\delta_d^+(x) \geqslant \delta_d^-(x). \tag{11}$$

Since $\widetilde{f}(y) = f(y)$ at the integer points, and since $\widetilde{f}$ is convex, we have

$$\delta_d^+(x) \leqslant \frac{f(x+d) - f(x)}{|d|}, \qquad \delta_d^-(x) \geqslant \frac{f(x) - f(x-ad)}{|d|},$$

where $|d|$ is the L_2 norm of d. This, together with (11) imply the integer convexity of f. ∎

The converse of the above theorem is not true:

Counter-example 2.1. Consider the convex function $f : \mathbb{N}^m \to \mathbb{R}$ given by $f(x) = \max_{i=1,\ldots,m} x_i$. It is integer convex since it is the maximum of convex (linear) functions. However, it is not multimodular. Indeed, consider $m = 2$, $x = (i+1, i)$ for some integer i. Then

$$2i + 2 = f(x - e_1 + e_2) + f(x) > f(x - e_1) + f(x + e_2) = 2i + 1.$$

Hence f is not multimodular.

Actually, when dealing with discrete functions, the counterpart of convexity is not integer convexity but rather multimodularity, which insures, for example, that a local minimum is a global minimum point (see Corollary 2.2).

The following result will be very useful when cost functions involve arbitrary convex functions of the quantities of interest, especially to prove optimality in the convex increasing order.

Theorem 2.3. *The following propositions are equivalent,*
(i) f be multimodular and for all $v, w \in \mathcal{F}$ and $\max(f(x+v), f(x+w)) \geqslant \max(f(x), f(x+v+w))$.
(ii) For all $h : \mathbb{R} \to \mathbb{R}$ convex increasing, $h(f)$ is multimodular.

Proof. We will use the notation introduced in [9]. Condition (i) is equivalent to the fact that the two-dimensional vector $(f(x), f(x+v+w))$ is weakly submajorized by $(f(x+v), f(x+w))$, denoted $(f(x), f(x+v+w)) \prec_w (f(x+v), f(x+w))$. By using proposition 4.C.1.b in [9] then this implies that for any convex increasing function h, $h(f(x)) + h(f(x+v+w)) \leqslant h((f(x+v)) + h(f(x+w))$. This is exactly the multimodularity of the function $h(f)$.

As for the converse, note that (ii) implies in particular that $h(f)$ is multimodular for all h continuous, convex and increasing. Using Proposition 4.B.2 in [9], this implies that $(f(x), f(x+v+w)) \prec_w (f(x+v), f(x+w))$. By definition of weak majorization, that is the same as statement (i). ∎

3 The optimality of balanced policies for a single criterion

In this section, we will present a rather general framework under which modularity can be used in order to optimize a cost function based on a sequence of functions which will represent a quantity of interest in a given model. The problems that can be solved using this framework include the minimization of the average workload in a queue under general stationary assumptions, as well as many other similar problems (see for example [1]). Section § 5.2 presents a precise instance of such a problem.

The sequence of functions considered can be interpreted as cost functions, which are defined on a common sequence $(a_1, a_2, ...)$ of integers which we call a *control sequence.* For example, a_i can be the number of admitted jobs at the ith arriving epoch for an admission control problem in a queue. The control sequences will all belong to a set A which is a convex union of atoms in $\mathbb{Z}^k$. Our objective is to study optimization properties of Cesaro averages of the cost functions over the class of control sequences.

Consider a sequence of functions $f_k : \mathbb{N}_+^k \to \mathbb{R}_+ \cup \{\infty\}$ that satisfy the following assumptions:

- $< 1 >$ f_k is multimodular on A.
- $< 2 >$ $f_k(a_1, ..., a_k) \geqslant f_{k-1}(a_2, ..., a_k), \ \forall k > 1;$

For a given sequence $\{a_k\}$, we define the cost $g(a)$ as

$$g(a) = \overline{\lim_{N\to\infty}} \frac{1}{N} \sum_{n=1}^{N} f_n(a_1, ..., a_n). \tag{12}$$

Definition 3.1. *Let p and θ be two positive reals. We define the balanced sequence $\{a_k^p(\theta)\}$ with rate p and initial phase θ as,*

$$a_k^p(\theta) = \lfloor kp + \theta \rfloor - \lfloor (k-1)p + \theta \rfloor, \tag{13}$$

where $\lfloor x \rfloor$ is the largest integer smaller than or equal to x.

Note that the set $\{a_k^p(\theta), 0 \leqslant \theta \leqslant 1\}$ are extreme points of an atom containing the point $(p, p, ..., p)$.

The aim of this section is to prove that this sequence minimizes the function g, provided that some conditions (including $< 1 >$ and $< 2 >$) above hold. The sequence (13) was used by Hajek in [6], and we use several properties of the balanced sequence established in [6]. To establish the main optimization results, we need the following technical Lemma.

Lemma 3.1. *If f_k satisfies assumption $< 2 >$, then the function $\widetilde{f_k}$ satisfies assumption $< 2 >$ for positive real numbers.*

The proof of this lemma is given in Appendix B.

Lemma 3.2. *Under assumptions $< 1 >$ and $< 2 >$, let Θ be a random variable, uniformly distributed in $[0,1]$, and denote the expectation w.r.t. Θ by E_Θ. Then*

$$\lim_{N\to\infty} E_\Theta f_N(a_1^p(\Theta),...,a_n^p(\Theta)) = \lim_{N\to\infty} \widetilde{f}_N(p,p,...,p). \tag{14}$$

Proof We have for all n,

$$E_\Theta f_n(a_1^p(\Theta),...,a_n^p(\Theta)) = \widetilde{f}_n(p,...,p). \tag{15}$$

(This follows (8), from Lemma 2.4, and the fact that $\widetilde{f}_n$ is affine on each atom, and agrees with f_n for the extreme points of the atom.) Since $\widetilde{f}_N(p,p,...,p)$ is increasing in N by Lemma 3.1, the limit in N exists (it is possibly infinite). ∎

Definition 3.2. *We call the sequence $\{a^p(\Theta)\}$ the randomized balanced policy with rate p.*

3.1 Upper Bounds

Lemma 3.3. *Under assumptions $< 1 >$ and $< 2 >$, for every $\theta \in [0,1]$,*

$$\overline{\lim_{N\to\infty}} \frac{1}{N}\sum_{n=1}^{N} f_n(a_1^p(\theta),...,a_n^p(\theta)) \leqslant \lim_{N\to\infty} \widetilde{f}_N(p,p,...,p). \tag{16}$$

Proof. Define

$$f_m(\theta,p) \triangleq f_m(a_1^p(\theta),...,a_m^p(\theta)).$$

f_m is periodic (in θ) with period 1. Define

$$f'_m(\theta,p) \triangleq f_m(a^p_{-m+1}(\theta),...,a_0^p(\theta)).$$

Then we have

$$f'_m(\theta',p) = f_m(\theta,p) \quad \text{where} \quad \theta' = \theta - mp, \tag{17}$$

Indeed,

$$\begin{aligned} f'_m(\theta',p) &= f_m(a^p_{-m+1}(\theta'),...,a_0^p(\theta')) \\ &= f_m(a^p_{-m+1}(\theta+mp),...,a_0^p(\theta+mp)) = f_m(\theta,p), \end{aligned}$$

where the last equality follows from the fact that $a^p_{-m+k}(\theta+mp) = a_k^p(\theta)$, $k=1,...,m$. f'_m is again periodic w.r.t. θ, with period 1, and is increasing in m so that the following limit exists (possibly infinite):

$$f'_\infty(\theta,p) \triangleq \lim_{m\to\infty} f'_m(\theta,p).$$

Moreover, we have that $E_\Theta f'_m(\Theta,p) = \widetilde{f}_m(p,...,p)$, where Θ be a random variable, uniformly distributed in $[0,1]$ (this follows from (8), from Lemma 2.4, and fact that $\widetilde{f}_n$ is affine on each atom, and agrees with f_n for the extreme points of the atom). Hence,

$$E_\Theta f'_\infty(\Theta,p) = \lim_{N\to\infty} \widetilde{f}_N(p,p,...,p). \tag{18}$$

Consider now the balanced sequence for fixed θ. Then

$$\begin{aligned}
&\frac{1}{N}\sum_{m=1}^{N} f_m(a_1^p(\theta),...,a_m^p(\theta)) \\
&\quad\leqslant \frac{1}{N}\sum_{m=1}^{N} f_N(a^p_{-N+m+1}(\theta),...,a_0^p(\theta),...,a_m^p(\theta)) \\
&\quad\leqslant \frac{1}{N}\sum_{m=1}^{N} f'_\infty(-mp+\theta,p).
\end{aligned}$$

The last inequality follows from assumption $< 2 >$ for the functions f_k, as well as an argument similar to the one used in (17).

If p is irrational, applying the ergodic theorem of Weyl and Von Neumann ([11]), we have

$$\lim_{N\to\infty}\frac{1}{N}\sum_{m=1}^{N} f'_\infty(-mp+\theta,p) = E_\Theta f'_\infty(\Theta,p).$$

From equation (18), we have $E_\Theta f'_\infty(\Theta,p) = \lim_{N\to\infty}\widetilde{f}_N(p,p,...,p)$. This implies that if p is irrational,

$$\lim_{N\to\infty}\frac{1}{N}\sum_{m=1}^{N} f'_\infty(-mp+\theta,p) = \lim_{N\to\infty}\widetilde{f}_N(p,p,...,p). \tag{19}$$

If p is rational, then $p = q/d$ were q and d are relatively prime and $d \geqslant 1$. This implies that the sequence $(a^p_{-(N-m)}(\theta),...,a_0^p(\theta),...,a_m^p(\theta))$ is constant if $\theta(\bmod 1) \in [j/d,(j+1)/d)$, for all j. Therefore, $f'_m(\theta,p)$ is also constant on these intervals and by passage to the limit, $f'_\infty(\theta,p)$ is constant on these intervals. Now, note that $\mathrm{Frac}(\theta - mp) \in [j/d,(j+1)/d)$ for exactly one value of m out of d consecutive values of m because q are d are relatively prime.

Now, we have

$$\lim_{N\to\infty}\frac{1}{N}\sum_{m=1}^{N} f'_\infty(-mp+\theta,p) = \frac{1}{d}\sum_{m=1}^{d-1} f'_\infty(m/d,p) = E_\Theta f'_\infty(\Theta,p).$$

Equation (18) concludes this case as well. ∎

3.2 Lower Bounds

In this subsection, we establish lower bounds for the discounted cost for all control sequences $\{a_k\}$. This then serves for obtaining a lower bound on the average cost. Here, we use the following assumption for the functions f_k.

- $< 3 >$ For any sequence $\{a_k\}$ $\exists$ a sequence $\{b_k\}$ such that $\forall k, m, \quad k > m, \; f_k(b_1, \cdots, b_{k-m}, a_1, ..., a_m) = f_m(a_1, ..., a_m)$.

We use the notions defined in the previous sections.

Let us fix the sequence $\{a_k\}$, as well as some arbitrary integer, N. We define $p_\alpha \triangleq (1-\alpha)\sum_{k=1}^{\infty} \alpha^{k-1} a_k$.

Now, using assumptions $< 1 >$ through $< 3 >$, we have

$$\begin{aligned}
&\sum_{n=1}^{\infty}(1-\alpha)\alpha^{n-1} f_n(a_1, a_2 ..., a_n) \\
&\geqslant \sum_{n=1}^{N}(1-\alpha)\alpha^{n-1} f_N(b_1, \cdots b_{N-n}, a_1, a_2 ..., a_n) + \sum_{n=N+1}^{\infty}(1-\alpha)\alpha^{n-1} f_N(a_{n-N+1}, a_{n-N+2} ..., a_n) \\
&= \sum_{n=1}^{N}(1-\alpha)\alpha^{n-1} \widetilde{f}_N(b_1, \cdots b_{N-n}, a_1, a_2 ..., a_n) + \sum_{n=N+1}^{\infty}(1-\alpha)\alpha^{n-1} \widetilde{f}_N(a_{n-N+1}, a_{n-N+2} ..., a_n) \\
&\geqslant \widetilde{f}_N\big(\sum_{n=1}^{N}(1-\alpha)\alpha^{n-1}(b_1, \cdots b_{N-n}, a_1, a_2 ..., a_n) \\
&\qquad + \sum_{n=N+1}^{\infty}(1-\alpha)\alpha^{n-1}(a_{n-N+1}, a_{n-N+2} ..., a_n)\big) \qquad (20) \\
&= \widetilde{f}_N\left(b_1 \sum_{n=1}^{N-1}(1-\alpha)\alpha^{n-1} + \alpha^N p_\alpha, b_2 \sum_{n=1}^{N-2}(1-\alpha)\alpha^{n-1} + \alpha^{N-1} p_\alpha, \cdots, p_\alpha\right). \qquad (21)
\end{aligned}$$

Equation (20) follows from Jensen's inequality, since by Theorem 2.1, the function $\widetilde{f}_N$ is convex, and since the coefficients $(1-\alpha)\alpha^{n-1}$ are nonnegative and sum to 1. Define

$$B(N, \alpha, p) \triangleq \widetilde{f}_N\left(b_1 \sum_{n=1}^{N-1}(1-\alpha)\alpha^{n-1} + \alpha^N p, b_2 \sum_{n=1}^{N-2}(1-\alpha)\alpha^{n-1} + \alpha^{N-1} p, \cdots, p\right). \qquad (22)$$

Note that B is defined for a fixed sequence $\{a_k\}$. Also note that $B(N, \alpha, p)$ is lower semi-continuous in α and in p.

Using Lemma C.1 in Appendix C, we derive the following lower bounds

Lemma 3.4. *Under assumptions* $< 1 >$, $< 2 >$ *and* $< 3 >$,

$$\overline{\lim_{m\to\infty}} \frac{1}{m}\sum_{n=1}^{m} f_n(a_1,\cdots,a_n) \geqslant \inf_{q\in\mathcal{L}} \widetilde{f}_N(q,...,q),$$

where $\mathcal{L}$ *is the set of all limit points of* p_α *as* $\alpha \uparrow 1$.

Proof.

$$\begin{aligned}\overline{\lim_{m\to\infty}} \frac{1}{m}\sum_{n=1}^{m} f_n(a_1,\cdots,a_n) &\geqslant \overline{\lim_{\alpha\uparrow 1}}(1-\alpha)\sum_{n-1}^{\infty}\alpha^{n-1}f_n(a_1,\cdots,a_n) \\ &\geqslant \overline{\lim_{\alpha\uparrow 1}} B(N,\alpha,p_\alpha) \\ &\geqslant \inf_{q\in\mathcal{L}}B(N,1,q), \end{aligned} \tag{23}$$

The Lemma follows since for any given p, by definition of B, $B(N,1,p) = \widetilde{f}_N(p,p,\cdots,p)$. ∎

3.3 Optimality of the Balanced Sequences

Theorem 3.1. *Under assumptions* $< 1 >$,$< 2 >$ *and* $< 3 >$, *and given some* $p \in [0,1]$, *and any* $\theta \in [0,1]$, *if the functions* $f_k(a_1,\cdots,a_k)$ *are increasing in all* a_i, *then the balanced sequence* $a^p(\theta)$ *minimizes the average cost* $g(a)$ *over all sequences that satisfy the constraint:*

$$\underline{\lim_{N\to\infty}} \frac{1}{N}\sum_{n=1}^{N} a_n \geqslant p.$$

Proof.
We denote by

$$\underline{p} \overset{\triangle}{=} \underline{\lim_{N\to\infty}} \frac{1}{N}\sum_{n=1}^{N} a_n$$

By using Lemma C.1 in the Appendix C,

$$p \leqslant \underline{p} \leqslant \underline{\lim_{\alpha\uparrow 1}} p_\alpha = \inf\{q, q\in\mathcal{L}\}.$$

If the functions $\{f_k\}$ are increasing, then B is increasing in p, therefore,

$$g(a) \geqslant \inf_{q\in\mathcal{L}}B(N,1,q) \geqslant B(N,1,p) = \widetilde{f}_N(p,...,p), \tag{24}$$

by Lemma 3.4. If we let N go to infinity, we get

$$g(a) = \overline{\lim_{m\to\infty}} \frac{1}{m}\sum_{n-1}^{m} f_n(a_1,\cdots,a_n) \geqslant \overline{\lim_{N\to\infty}} \widetilde{f}_N(p,\cdots,p). \tag{25}$$

Lemma 3.3 shows that $\overline{\lim}_{N\to\infty} \widetilde{f}_N(p,p,p) \geqslant g(a^p(\theta))$. Thus $g(a) \geqslant g(a^p(\theta))$. ∎

When the functions f_k are decreasing, we have the analogous result.

Theorem 3.2. *Under assumptions $< 1 >$,$< 2 >$ and $< 3 >$, and given some $p \in [0,1]$, and any $\theta \in [0,1]$, if the functions $f_k(a_1, \cdots, a_k)$ are decreasing in all a_i, then the balanced sequence $a^p(\theta)$ minimizes the average cost $g(a)$ over all sequences that satisfy the constraint:*

$$\overline{\lim_{N\to\infty}} \frac{1}{N} \sum_{n=1}^{N} a_n \leqslant p.$$

Proof. The proof is similar to the previous one, using the fact that if

$$\overline{p} \stackrel{\triangle}{=} \overline{\lim_{N\to\infty}} \frac{1}{N} \sum_{n=1}^{N} a_n,$$

then

$$p \geqslant \overline{p} \geqslant \overline{\lim_{\alpha\uparrow 1}} p_\alpha = \sup\{q, q \in \mathcal{L}\}.$$

■

Remark 3.1. In Theorem 5.2 of [6], Hajek proves the optimality of the sequence $a^p(\theta)$ for any cost function of the form

$$h(a) \stackrel{\triangle}{=} \underline{\lim_{N\to\infty}} \frac{1}{N} \sum_{n=m}^{N} f(a_{n-m+1}, a_{n-m+2}, ...a_n), \tag{26}$$

where f is a multimodular function with m variables.
In the rest of his paper, he applies this result to minimize the average cost $g(a)$ of an admission sequence a in a GI/M/1 queue. However, $g(a)$ is not of the form given in Equation (26), because each immediate cost depends on the whole sequence $a_1 \cdots a_n$, rather than on the last m decisions. In fact, the average cost is of the form given in Equation (12). Therefore, in order to show the optimality of the balanced policy for the average cost $g(a)$, a rather extensive analysis, using intricate stability arguments, is needed in section 6 of [6].

In our derivation the form of the cost function we are using involves a sequence $\{f_n\}$, where f_n has n arguments (rather than a fixed number m, independent of n, as is the case in [6]). This, together with the lower bounds obtained via Abel limits, allows us to avoid any stability arguments, because this form of the cost function is adapted to average costs in queues. It allows us to apply Theorems 3.1 and 3.2 directly to queuing systems. Also, this approach is more general. In fact we can study (max,+) linear networks (see [1]) instead of single queues, under mere stationary assumptions (independence is not needed).

Moreover, in the next section our approach gives optimality of a balanced policy for the average cost of the routing to $K = 2$ queues, and also for many special cases with $K \geqslant 3$ queues (see [2]).

4 The optimality of balanced policies for multiple criteria

In this section, we establish general conditions under which the balanced policy is optimal when the cost function depends on multiple criteria. While the single criterion framework has application in admission control, this multiple criteria approach has applications for routing control to several queues. For instance, for the routing to several identical ./GI/1 queues, it is known that the round robin routing is optimal in separable-convex increasing order [8]. In order to obtain this type of results in our framework, we will consider a cost criterion of the form (27). It turns out that we can establish lower bounds (Theorem 4.1 for a general convex function h. Moreover, we show the existence of optimal asymptotic fractions in a very genera; case. As for the optimality proof of the balanced policy, (if it exists), we have restricted ourselves to linear functions in Theorem 4.2.

From now on, we study the following general optimization problem. Consider K sequences of functions $f_n^{(i)}, i = 1, ..., K$. Each sequence of functions $f_n^{(i)}$ will only depend on the ith coordinates a^i in a, and will satisfy assumptions $< 1 >$,$< 2 >$ and $< 3 >$, as in Section 3.

A policy is a sequence $a = (a_1, a_2, ...)$, where a_n is a vector taking values in $\{0,1\}^K$. We consider the additional constraint that for every integer j, only one of the components of a_j may be different than 0. A policy satisfying this constraint is called feasible.

Let h be a convex increasing function from $\mathbb{R}^K$ to $\mathbb{R}$. Define

$$g(a) \stackrel{\triangle}{=} \overline{\lim_{N\to\infty}} \frac{1}{N} \sum_{n=1}^{N} h(f_n^1(a), ..., f_n^K(a)), \tag{27}$$

Following notations introduced in Section 3, we get a bounding function called $B_i(N, \alpha, p)$ for coordinate i. Here, we denote by

$$B_i(\alpha, p) \stackrel{\triangle}{=} \sup_N B_i(N, \alpha, p),$$

and

$$B_i(p) \stackrel{\triangle}{=} \sup_{\alpha \leqslant 1} B_i(\alpha, p).$$

Note that by convexity of $\widetilde{f}_n^{(i)}$ and Lemma 3.1, $B_i(\alpha, p)$ and $B_i(p)$ are continuous from below in (α, p) and p, respectively.

Our objective is to minimize $g(a)$, with no constraints on the asymptotic fractions.

Theorem 4.1. *Assume that for all i, the functions $f_n^{(i)}$ satisfy assumptions $< 1 >$,$< 2 >$ and $< 3 >$. The following lower bound holds for all policies:*

$$g(a) \geqslant \inf_{p_1+...+p_K=1} h(B_1(p_1), ..., B_K(p_K)).$$

Proof. Due to Lemma C.1 in the Appendix C, Jensen's inequality and equation (21), we have

$$\begin{aligned}
&\overline{\lim_{N\to\infty}} \frac{1}{N}\sum_{n=1}^{N} h(f_n^1, ..., f_n^K) \\
\geqslant\ & \overline{\lim_{\alpha\to 1}}(1-\alpha)\sum_{n=1}^{\infty}\alpha^{n-1}h(f_n^1, ..., f_n^K) \\
\geqslant\ & \overline{\lim_{\alpha\to 1}}\, h\left((1-\alpha)\sum_{n=1}^{\infty}\alpha^{n-1}f_n^1, ..., (1-\alpha)\sum_{n=1}^{\infty}\alpha^{n-1}f_n^K\right) \\
\geqslant\ & \overline{\lim_{\alpha\to 1}}\, h\left(B_1(\alpha, p_1^a(\alpha)), ..., B_K(\alpha, p_K^a(\alpha))\right), \qquad (28)
\end{aligned}$$

where

$$p_i^a(\alpha) \triangleq (1-\alpha)\sum_{k=1}^{\infty}\alpha^{k-1}a_k^i. \qquad (29)$$

We note that $\sum_{i=1}^{K} p_i^a(\alpha) = 1$. Hence, one may choose a sequence $\alpha_n \uparrow 1$ such that the following limits exist:

$$\lim_{n\to\infty} p_i^a(\alpha_n) \triangleq p_i, \qquad i = 1, ..., K \qquad (30)$$

and $\sum_{i=1}^{K} p_i = 1$. From the lower semi-continuity of $B_i(\alpha, p_i)$ in p_i and α we get from (28)

$$\begin{aligned}
g(a) \ &\geqslant\ h(B_1(1,p_1), ..., , B_K(1,p_K)) \qquad (31)\\
&\geqslant\ \inf_{p_1+...+p_K=1} h(B_1(1,p_1), ..., B_K(1,p_K)).
\end{aligned}$$

■

Note that there exists some p^* that achieves the infimum

$$\inf_{p_1+...+p_K=1} h(B_1(1,p_1), ..., B_K(1,p_K)), \qquad (32)$$

since $h(B_1(1,p_1), ..., B_K(1,p_K))$ is continuous from below in $p = (p_1, \cdots, p_K)$.

Consider the policy $a^{p^*}(\theta)$ given by

$$a_{k,i}^{p^*}(\theta) = \lfloor kp_i^* + \theta_i \rfloor - \lfloor (k-1)p_i^* + \theta_i \rfloor. \qquad (33)$$

There are some p^* for which the condition of feasibility of the policy $a^{p^*}(\theta)$ is satisfied, that is, there exists some $\theta = (\theta_1, ..., \theta_K)$, such that the policy $a^{p^*}(\theta)$ given in (33) is feasible. These p^* are called *balanced* and are more exhaustively studied in [2] and references therein.

Theorem 4.2. *Assume that h is linear increasing and that p^* is balanced. Then $a^{p^*}(\theta)$ is optimal for the average cost, i.e. it minimizes $g(a)$ over all feasible policies.*

Proof. The proof follows directly from Lemma 3.3 together with Theorem 4.1. ∎

The balance condition on p^* is still not completely characterized, however, we can mention two simple cases for which p^* is balanced. i.e. for which there exist some $\theta = (\theta_1, ..., \theta_K)$, such that $a^{p^*}(\theta)$ is feasible.

- **P1:** $K = 2$.
- **P2:** K criteria with symmetric costs, *i.e.* $h(x) = \sum_i x_i$ and all f^i (as functions of a^i) are equal.

Corollary 4.1. *(i) Consider problem P1. For p^* that achieves the minimum in (32), the balanced policy for rate p^* and some initial phase θ is feasible and optimal.*
(ii) Consider problem P2. By symmetry, the balanced policy with $p = 1/K$ for some initial phase θ is feasible and optimal.

Remark 4.1. *Hajek gives an argument ([6] Remark (5) p. 554) that shows for $K = 2$ the optimality of the balanced policy with rate vector $(p, 1-p)$ in the restricted class of policies a such that $\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} a_n^1 = p$. Note that Corollary 4.1 (ii) shows the optimality of a balanced policy for all control sequences a.*

Next, we restrict again to the case of a single objective ($K = 1$), and show that the results of the previous section can be extended. More precisely, we show that balanced policy is optimal in a stronger sense.

Corollary 4.2. *Under the conditions of Theorem 3.1, given some $p \in [0, 1]$, and any $\theta \in [0, 1]$, the balanced policy $a^p(\theta)$ minimizes the average cost $g(a)$ over all policies that satisfy the constraint:*

$$\overline{\lim_{\alpha\to 1}}\, p_1(\alpha) \geqslant p_1. \tag{34}$$

where $p_1^a(\alpha)$ is defined in (29).

Note that the constraint $\underline{\lim}_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} a_n \geqslant p_1$ (in Theorem 3.1) implies (34), due to Lemma C.1 in the Appendix C. Therefore the minimization in Theorem 3.1 is over a subclass of the set of policies on which minimization is performed in Corollary 4.2. Thus, Corollary 4.2 implies that a policy a that satisfies (34) does not perform better than the balanced policy (with $p = p_1$) even if $\underline{\lim}_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} a_n < p_1$.

Proof of Corollary 4.2: Choose an arbitrary policy a that satisfies (34). Choose a subsequence $\alpha_n \uparrow 1$ such that $\overline{\lim}_{n\to\infty} p_1^a(\alpha_n) \geqslant p_1$. The proof now follows by combining Lemma 3.3 with (34). ∎

5 Application of the optimization theorems

In this section, we will briefly present some typical optimization problems that fit the framework presented in the previous sections. Other models that require a more extensive analysis can be found in [1, 2, 3].

5.1 Admission control to a queue

In this section we give a typical application of the multimodularity theory by looking at a G/G/1 queue with batch arrivals and an admission control a that must accept a given asymptotic proportion p of the arriving customers. A more detailed analysis of this system can be found in [1]. The model is illustrated in Figure 2.

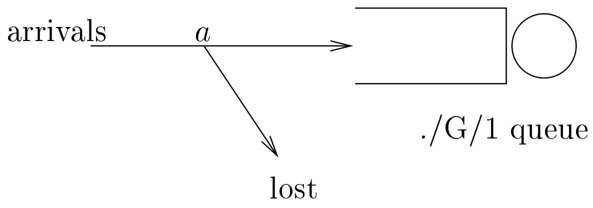

Figure 2: Admission control to a ./G/1 queue.

Let $\{T_i\}_{i\in\mathbb{N}}$ be the sequence of arrival times, with the convention that $T_1 = 0$, the queue being empty at time 0. The admission control is defined through an *control sequence*. The control sequence is a sequence of integer numbers, $a = (a_1, a_2, \cdots, a_N, \cdots)$, where a_i gives the number of customers admitted to the queue at time T_i. We also denote by $\{\sigma_k\}$ the sequence of service times in the server.

We denote by $W_k(a_1, \cdots, a_k)$ the workload in the queue at time T_k under the admission control a. Here, W_k will play the role of the functions f_k.

The following result is proved in [1] (the special case of a D/D/1 queue is analyzed in details in the next section). Let $\mathbf{E}_{\sigma,T}$ denote the expectation w.r.t. the service times and the inter-arrival times.

Theorem 5.1. *Assume that $\{T_k\}$ and $\{\sigma_k\}$ are stochastically independent stationary sequences. Then for all $k, m \in \mathbb{N}$ the functions $\mathbf{E}_{\sigma,T}W_k(a_1, \cdots, a_k)$ have the following properties:*

- $\mathbf{E}_{\sigma,T}W_k(a_1, \cdots, a_k)$ *is multimodular on the restriction to* $\mathbb{N}^k$.
- $\mathbf{E}_{\sigma,T}W_k(a_1, \cdots, a_k) \geqslant \mathbf{E}_{\sigma,T}W_m(a_{k-m+1}, \cdots, a_k)$, *for* $k > m$.
- $\mathbf{E}_{\sigma,T}W_k(a_1, \cdots, a_k) = \mathbf{E}_{\sigma,T}W_m(0, \cdots, 0, a_1, \cdots, a_k)$, *for* $k < m$.
- $\mathbf{E}_{\sigma,T}W_k(a_1, \cdots, a_k)$ *is increasing in* a_i, $i = 1, ..., k$.

Remark 5.1. *It seems difficult to give a consistent meaning to the workload associated with a negative number of customers. One may even doubt that there is a satisfying way to do so that will also preserve multimodularity. This is why we have not considered the quantity $\mathbf{E}_{\sigma,T}W_k(a_1, \cdots, a_k)$ on $\mathbb{Z}^k$, but only on $\mathbb{N}^k$, which is a convex union of atoms. Therefore, all the optimization framework constructed in section 3 can be used in this case.*

The expectation of (any increasing convex function h of) the workload also satisfies conditions $< 1 >$,$< 2 >$ and $< 3 >$. The general theorem 3.1 applies and the balanced admission policy $a^p(\theta)$ with rate p minimizes

$$g(a) \stackrel{\triangle}{=} \overline{\lim_{N\to\infty}} \frac{1}{N} \sum_{n=1}^{N} \mathbf{E}_{\sigma,T} W_n(a_1, ..., a_n),$$

among all policies with rate at least p.

In [1] we also consider the problem of minimizing average waiting times and therefore we need Theorem 3.2.

This G/G/1 model can be generalized. In [1], the traveling time of a customer in an arbitrary network of queues which forms an event graph is shown to be multimodular with respect to the admission sequence. The optimality of the balanced sequence in this case is proved in [1]. More general applications of these results can also be found in [2].

5.2 Applications in high-speed telecommunication systems

In this section we present another illustration of the theorems that are given in sections 3 and 4 that we fully develop. Here, we consider a simple model composed of a controlled D/D/1 queue with service times $\sigma_n = \sigma$ and inter-arrival times $\tau_n = \tau$ all deterministic. Assume that the available actions are 0 (corresponding to rejecting an arriving customer) and 1 (corresponding to acceptance of an arriving customer).

The type of problem we consider is typical in high speed telecommunications networks, and in particular, to the ATM (Asynchronous Transfer Mode). The latter has been chosen by the standardization committee ITU-T [7] as the main standard for integration of services in broadband networks. In order to handle efficiently a large variety of applications, such as voice, data, video and file transfer, cells of fixed size are used, giving rise to our model that uses fixed service times. Fixed inter-arrival times are typical for isochronous applications (voice, video) and also for large file transfer.

Two important measures of quality of services in ATM networks are loss probabilities (CLR - Cell Loss Ratios) and delays. According to the ATM standard [7], when a CBR (Constant Bit Rate) session is established, the network should provide a guarantee that these two measures are bounded by given constants. Since the available sources are limited and, moreover, might be shared with other applications, a typical objective of the network is to minimize the delay of the CBR session while meeting the constraint on the loss probabilities. Losses might be due either to overflow, or to deliberate packet discarding by the network (e.g. to allow the resources to be available for other applications). The problem can be formulated in our framework as one of discarding cells so as to minimize the average queue size (i.e. the workload in the system) which is known to be proportional to the average sojourn time (due to Little's law), subject to a lower bound p on the average cell discarding rate.

We now describe the state evolution of the system. If x_n denotes the amount of workload in the system immediately after the nth arrival that occurs after time 0, and the system is initially empty (at time 0), then

$$x_n = \max(x_{n-1} - \tau, 0) + a_n\sigma.$$

The solution of this recursion is given by the expansion of the Lindley equation:

$$x_n = f_n(a_1, ..., a_n) = \max\left\{0, \sum_{k=j}^{n-1}(a_k\sigma - \tau), j = 1, ..., n-1\right\} + a_n\sigma. \tag{35}$$

We show by a simple inductive argument that for all m, x_m satisfies:

$$\begin{aligned} x_n(a+v_1) + x_n(a+v_2) &\geqslant x_n(a) + x_n(a+v_2+v_1), & (36)\\ x_n(a+v_1) \vee x_n(a+v_2) &\geqslant x_n(a) \vee x_n(a+v_2+v_1). & (37)\end{aligned}$$

The function $x_1(a) = a_1\sigma$ clearly satisfies 36 and 37.

if v_1, v_2 are in $\mathcal{F}\backslash\{e_n, d_n\}$, then by induction

$$\begin{aligned} &x_n(a+v_1) + x_n(a+v_2) = (x_{n-1}(a+v_1) - \tau \vee 0) + a_n\sigma + (x_{n-1}(a+v_2) - \tau \vee 0) + a_n\sigma \\ &= (x_{n-1}(a+v_1) + x_{n-1}(a+v_2) - 2\tau \vee x_{n-1}(a+v_1) - \tau \vee x_{n-1}(a+v_2) - \tau \vee 0) + 2a_n\sigma \\ &\geqslant (x_{n-1}(a) + x_{n-1}(a+v_2+v_1) - 2\tau \vee x_{n-1}(a) - \tau \vee x_{n-1}(a+v_2+v_1) - \tau \vee 0) + 2a_n\sigma \\ &= x_n(a) + x_n(a+v_2+v_1).\end{aligned}$$

$$\begin{aligned} x_n(a+v_1) \vee x_n(a+v_2) &= (x_{n-1}(a+v_1) - \tau \vee x_{n-1}(a+v_2) - \tau \vee 0) + a_n\sigma \\ &= ((x_{n-1}(a+v_1) \vee x_{n-1}(a+v_2)) - \tau \vee 0) + a_n\sigma \\ &\geqslant x_n(a) \vee x_n(a+v_2+v_1).\end{aligned}$$

If $v_1 = d_m$ and v_2 is in $\mathcal{F}\backslash\{e_n, d_n\}$, then equation 36 is obtained similarly, by induction. As for equation 37, the proof is slightly different.

$$\begin{aligned} x_n(a+v_1) \vee x_n(a+v_2) &= (x_{n-1}(a+e_{n-1}) - \tau - \sigma \vee -\sigma \vee x_{n-1}(a+v_2) - \tau \vee 0) + a_n\sigma \\ &= (x_{n-1}(a) - \tau \vee x_{n-1}(a+v_2) - \tau \vee 0) + a_n\sigma \\ &\geqslant x_n(a) \vee x_n(a+v_2+v_1).\end{aligned}$$

If $v_1 = e_m$, $x_n(a+e_m) + x_n(a+v_2) = x_n(a) + x_n(a+v_2) + \sigma = x_n(a) + x_n(a+v_2+e_m)$, and

$$\begin{aligned} x_n(a+e_m) \vee x_n(a+v_2) &= x_n(a) + \sigma \\ &\geqslant x_n(a) \vee x_n(a+v_2) + \sigma = x_n(a) \vee x_n(a+v_2+v_1).\end{aligned}$$

Our goal is now to obtain a policy a^* that minimizes an expected average cost related to the amount of work in the system at arrival epochs. The cost to be minimized is thus

$$g(a) \triangleq \overline{\lim_{N\to\infty}} \frac{1}{N} \sum_{n=1}^{N} f_n(a_1, ..., a_n),$$

subject to the constraint:

$$\varliminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} a_n \geqslant p^*.$$

Consider first the case of a queue with infinite capacity. Then, it follows from Theorem 3.1 that a balanced policy with rate p^* and arbitrary θ is optimal. The assumptions of the Theorem indeed hold:

- f_n (in (35)) is indeed monotone increasing in a_i;
- Property $< 3 >$ (in Subsection 3.2) holds by choosing $b_k = 0$, since

$$f_k(a_1, ..., a_k) = f_m(\underbrace{0, ..., 0}_{m-k}, a_1, ..., a_k), \;\; k < m; \tag{38}$$

- By combining (38) with the first monotonicity property, we get

$$f_{k-1}(a_2, ..., a_k) = f_k(0, a_2, ..., a_k) \leqslant f_k(a_1, ..., a_k),$$

  which establishes Property $< 2 >$ (in the beginning of Section 3).

Consider now a queue with a finite storage capacity for the workload, i.e. the workload at the queue at each time instant is bounded by C. When the queue is full, the overflow workload is lost. The balanced policy need not be optimal anymore, as the following example shows.

Counter-example 5.1. *(Non optimality of a balanced policy)*
Let $\tau = 1$, $\sigma = 100$, $C = 100$, $p^* = 0.01$. Assume that the cost to be minimized is the average queue length. The balanced policy with rate 0.01 achieves an average queue length of 50.5 for any θ. Consider now the periodic policy of period 200 that accepts 2 consecutive customers and rejects all following ones. After the second acceptance, the amount of work in the system is 100 due to the limit on the queue capacity, and there is loss of workload (of 99 units). The average queue length is 25.75. Thus the new policy achieves half the queue length as the previous one.

■

Although the balanced policy in the above counter example results in a larger queue, it has the advantage over the other policy of not creating losses. As we now show, a balanced policy is optimal if we restrict to policies with the further constraint that no losses are allowed. Thus, consider the class of policies that satisfy the constraint:

$$x_n(a_1, ..., a_n) \leqslant C$$

where x_n is given by (35).

Lemma 5.1. *The set* $A \triangleq \{(a_1, \cdots, a_n) \in \mathbb{N}^n : x_n(a_1 \cdots, a_n) \leqslant C\}$ *defines a convex union of atoms of* $\mathbb{Z}^n$.

Proof. By definition of x_n, the set A can also be defined as

$$A = \{(a_1, \cdots, a_n) : \forall j, \quad a_j \in \mathbb{N}, \quad a_j \geqslant 0, \quad \sum_{k=j}^{n} a_k \leqslant \frac{C + (n-j)\tau}{\sigma}\}. \tag{39}$$

Since A is only formed of integer points, we can also write

$$A = \{(a_1, \cdots, a_n) : \forall j, \quad a_j \in \mathbb{N}, \quad a_j \geqslant 0, \quad \sum_{k=j}^{n} a_k \leqslant \lfloor \frac{C + (n-j)\tau}{\sigma} \rfloor\}. \tag{40}$$

Now, let us consider the constraints one by one.

- The constraints $a_j \geqslant 0$ restrict A to N^n which is made of a convex union of atoms.
- The constraint $a_n \leqslant \lfloor \frac{C}{\tau} \rfloor$ also restricts A on a union of atoms.
- Now, let us look at a general constraint $\sum_{k=j}^{n} a_k \leqslant \lfloor \frac{C+(n-j)\tau}{\sigma} \rfloor$. On the projection over the last $n-j$ coordinates, this constraint is a convex union of faces of atoms. Therefore, on the whole set, this constraint is a union of atoms.

To finish the proof, remark that the intersection of convex union of atoms is a convex union of atoms. ■

Using now Corollary 2.1 and Theorem 3.1, we conclude that a balanced policy is again optimal.

In the above admission control we considered only the possibility of accepting or rejecting the whole arriving batch (of 100). In practice, arriving batches may correspond to cells originating from different sources, and it is often possible to reject only a part of the batch.

Assume, thus, that the available actions are $a \in \{0, 1, ..., \overline{N}\}$, where $a = i$ means accepting $i(100/\overline{N})$ units of workload. Assume that the batch size of 100 is an integer multiple of $\overline{N}$. We can thus split an arrival batch and accept only a fraction of it; more precisely, we can either reject it, or accept $1/\overline{N}$th of the batch, or $2/\overline{N}$th, etc.... The smallest unit of batch which we can accept (i.e. $\overline{N}^{-1}$) is called a mini batch.

Consider now the balanced policy $a^*[\overline{N}]$ that is given in (13) corresponding to $p = p^*\overline{N}$. In other words, instead of considering a target fraction p of the whole batch to be accepted, which is smaller than (or equal to) 1, the new target corresponds to the average number of mini batches to be accepted, and can be any real number between 0 and $\overline{N}$ (in particular, $p = \overline{N}$ will correspond to accepting $\overline{N}$ mini-batches, i.e. the whole original batch).

We may repeat the above calculation and show that this policy is optimal for the cases of (i) the infinite queue and (ii) the bounded queue, restricting to policies that do not generate losses. Moreover, for both cases, this policy is better than the one that consists of accepting or rejecting the whole batch according to the policy a^* defined above, since a^* is a feasible policy in our new problem, for which $a^*[\overline{N}]$ is optimal (Theorem 3.1).

In order to illustrate the last point, consider $\overline{N} = 10$. A balanced policy corresponds to acceptance of a mini-batch of 10 units, once every 10 time slots. The average queue length obtained by that policy is 5.5, i.e. about ten times less than the one obtained when the whole batch was to be accepted or rejected.

A Appendix A: proof of Lemma 2.2

Proof. (a) We shall show that for any $w, v \in \mathcal{F}$, there is a one to one map $z : \mathbb{Z}^m \to \mathbb{Z}^m$

$$\Delta_v \Delta_w f(x) = f(z(x)) - f(z(x) + v + w) - f(z(x) + v) - f(z(x) + w). \tag{41}$$

Hence the multimodularity implies $\Delta_v \Delta_w f \leqslant 0$. Since the map is one to one, we get also the converse. In the next four equations we illustrate (41) which establishes the proof.

Consider first $w = d_i, v = d_j$ $(v \neq w)$. Then

$$\begin{aligned} \Delta_v \Delta_w f(x) &= (\Delta_{i-1} - \Delta_i)(\Delta_{j-1} - \Delta_j) f(x) \\ &= (\Delta_{i-1} - \Delta_i)(f(x + e_{j-1}) - f(x + e_j)) \\ &= f(x + e_{j-1} + e_{i-1}) - f(x + e_{j-1} + e_i) - f(x + e_j + e_{i-1}) + f(x + e_j + e_i) \\ &= f(z + d_j + d_i) - f(z + d_j) - f(z + d_i) + f(z) \end{aligned} \tag{42}$$

where $z \stackrel{\triangle}{=} x + e_j + e_i$.

Let $v = e_m, w = -e_1$. Then

$$\begin{aligned} \Delta_v \Delta_w f(x) &= \Delta_m (f(x - e_1) - f(x)) \\ &= f(x + e_m - e_1) - f(x - e_1) - f(x + e_m) + f(x) \end{aligned} \tag{43}$$

Let $v = e_m, w = d_j$.

$$\begin{aligned} \Delta_v \Delta_w f(x) &= \Delta_m (\Delta_{j-1} - \Delta_j) f(x) \\ &= \Delta_m (f(x + e_{j-1}) - f(x + e_j)) \\ &= f(x + e_{j-1} + e_m) - f(x + e_{j-1}) - f(x + e_j + e_m) + f(x + e_j) \\ &= f(z + d_j + e_m) - f(z + d_j) - f(z + e_m) + f(z) \end{aligned} \tag{44}$$

where $z = x + e_j$.

Let $v = -e_1, w = d_j$.

$$\begin{aligned} \Delta_v \Delta_w f(x) &= \Delta_v (\Delta_{j-1} - \Delta_j) f(x) \\ &= \Delta_{-e_1} (f(x + e_{j-1}) - f(x + e_j)) \\ &= f(x + e_{j-1} - e_1) - f(x + e_{j-1}) - f(x + e_j - e_1) - f(x + e_j) \\ &= f(z + d_j - e_1) + f(z + d_j) + f(z - e_1) + f(z) \end{aligned} \tag{45}$$

where $z = x + e_j$.

(b.i) For any $w \in \mathcal{F}$,

$$\Delta_{e_1}\Delta_w f(x) = \Delta_w\Delta_{e_1} f(x) = -\Delta_w\Delta_{-e_1} f(z)$$

where $z = x + e_1$. The result follows from Lemma 2.2 (a).

(b.ii) Without loss of generality, assume that $i \leqslant j$. Then

$$\Delta_i\Delta_j f = \left(\Delta_{e_1} - \sum_{k=2}^{i}\Delta_{d_k}\right)\left(\sum_{l=j+1}^{m}\Delta_{d_l} + \Delta_{e_m}\right) f$$

The proof of (b.ii) is established by applying Lemma 2.2 (a) and Lemma 2.2 (b.i).

For $i < j$ we have

$$(\Delta_j - \Delta_i)\Delta_j f \;=\; \left(-\sum_{k=i+1}^{j}\Delta_{d_k}\right)\left(\sum_{l=j+1}^{m}\Delta_{d_l} + \Delta_{e_m}\right) f$$

and (b.iii) is established by applying Lemma 2.2 (a). For $i > j$ we have

$$\Delta_j = \Delta_i - \sum_{k=i+1}^{m-1}\Delta_{d_k} - \Delta_m + \Delta_{e_1} - \sum_{k=2}^{j}\Delta_k.$$

Hence

$$\Delta_j(\Delta_j - \Delta_i)f \;=\; \left(\Delta_{e_1} - \sum_{k=2}^{j}\Delta_{d_k}\right)\left(-\sum_{k=i+1}^{m}\Delta_{d_k} - \Delta_m + \Delta_{e_1} - \sum_{k=2}^{j}\Delta_k\right) f.$$

Again, (b.iii) is established by applying Lemma 2.2 (a) and 2.2 (b.i).
By taking $i = j - 1$ in Lemma 2.2 (b.iii), we obtain (b.iv).
(b.v) For $j < i$,

$$\Delta_{e_i}\Delta_{d_j} f = \left(\sum_{k=i}^{m}\Delta_{d_k} + \Delta_{e_m}\right)\Delta_{d_j} f.$$

For $j > i$,

$$\Delta_{e_i}\Delta_{d_j} f = \left(\Delta_{e_1} - \sum_{k=2}^{i}\Delta_{d_k}\right)\Delta_{d_j} f.$$

For both cases, the proof is established by applying Lemma 2.2 (a), and in the second case we use also Lemma 2.2 (b.i).
(b.vi)

$$\Delta_{d_1}\Delta_{d_i} f = (\Delta_{e_m} - \Delta_{e_1})\Delta_{d_i} f$$

The proof is established by applying Lemma 2.2 (a) and 2.2 (b.i).
(b.vii)

$$\Delta_{d_i}\Delta_{d_i} f = \Delta_{d_i}\left(-\sum_{j\neq i}\Delta_{d_j}\right) f \tag{46}$$

For $i \neq 1$, the proof is established by applying Lemma 2.2 (a) and 2.2 (b.i), where we replace in the summation Δ_{d_1} by $\Delta_{e_m} - \Delta_{e_1}$.
For $i = 1$ it follows from part (b.vi) of this Lemma and setting $i = 1$ in (46).
(c) $\Delta_{-e_1}\Delta_{e_2} \leqslant 0$ due to (4); $\Delta_{-e_1}\Delta_{d_2} = \Delta_{-e_1}(\Delta_{e_1} - \Delta_{e_2}) \leqslant 0$, and $\Delta_{e_2}\Delta_{d_2} = \Delta_{e_2}(\Delta_{e_1} - \Delta_{e_2}) \leqslant 0$ due (5). Hence f is multimodular by Lemma 2.2 (a). ∎

B Appendix B: Proof of Lemma 3.1

Let $z = (z_1, \cdots, z_k) \in \mathbb{R}^k_+$. This point belongs to an atom, say $S(z)$, made by the extreme points $x^0, x^1, \cdots, x^k$. The numbering of the extreme points of the atom is chosen such that according to the base $\mathcal{F}^k = (-e^k_1, d^k_1, \cdots, d^k_{k-1}, e^k_k)$, $x^1 = x^0 - e^k_1$. The other indices are arbitrary. This implies that $x^0_j = x^1_j$ for all $j > 2$. If we call P the projection of $\mathbb{R}^k_+$ onto $\mathbb{R}^{k-1}_+$ along the first coordinate,

$$\begin{aligned}
P(d^k_j) &= P(d^{k-1}_j) \text{ if } 2 < j < k \\
P(d^k_2) &= -e^{k-1}_1 \\
P(e^k_1) &= 0 \\
P(e^k_k) &= e^{k-1}_k \\
P(x^i) &= (x^i_2, \cdots, x^i_k)
\end{aligned}$$

These equalities imply that $P(x^0) = P(x^1)$ and $P(x^1), \cdots, P(x^k)$ form an atom in $\mathbb{R}^{k-1}_+$, as follows from the definition of an atom, and $P(z)$ belongs to this atom. Let

$$(z_1, z_2, \cdots, z_k) = \left(1 - \sum_{i=1}^{k}\alpha_i\right) x^0 + \alpha_1 x^1 + \cdots + \alpha_k x^k,$$

then

$$\begin{aligned}
(z_2, \cdots, z_k) &= P(z_1, z_2, \cdots, z_k) \\
&= \left(1 - \sum_{i=1}^{k}\alpha_i\right) P(x^0) + \alpha_1 P(x^1) + \cdots + \alpha_k P(x^k) \\
&= \left(1 - \sum_{i=2}^{k}\alpha_i\right) P(x^1) + \cdots + \alpha_k P(x^k).
\end{aligned}$$

Now,

$$\begin{aligned}
\widetilde{f}_k(z_1, z_2, \cdots, z_k) &= \left(1 - \sum_{i=1}^{k} \alpha_i\right) f_k(x^0) + \alpha_1 f_k(x^1) + \cdots + \alpha_k f_k(x^k) \\
&\geqslant \left(1 - \sum_{i=1}^{k} \alpha_i\right) f_{k-1}(P(x^0)) + \alpha_1 f_{k-1}(P(x^1)) + \cdots + \alpha_k f_{k-1}(P(x^k)) \\
&= \left(1 - \sum_{i=2}^{k} \alpha_i\right) f_{k-1}(P(x^1)) + \cdots + \alpha_k f_{k-1}(P(x^k)) \\
&= \widetilde{f}_{k-1}(z_2, \cdots, z_k).
\end{aligned}$$

C Appendix C: average and weighted costs

The following Lemma is often used in applications of optimal control (or games) with an average cost criteria (see e.g. [12]), yet it is not easy to find its proof in the literature in the format in which we want to apply it.

Lemma C.1. *Consider a sequence a_n of real numbers all having the same sign. Then,*

$$\overline{\lim_{n\to\infty}} \frac{1}{N} \sum_{n=1}^{N} a_n \geqslant \overline{\lim_{\alpha\to 1}} (1-\alpha) \sum_{k=1}^{\infty} \alpha^{k-1} a_k \tag{47}$$

$$\geqslant \underline{\lim}_{\alpha\to 1} (1-\alpha) \sum_{k=1}^{\infty} \alpha^{k-1} a_k \geqslant \underline{\lim}_{n\to\infty} \frac{1}{N} \sum_{n=1}^{N} a_n \triangleq p \tag{48}$$

Proof. Note that

$$\frac{1}{1-\alpha} \sum_{k=1}^{\infty} \alpha^{k-1} a_k = \sum_{k=1}^{\infty} \left(\sum_{l=1}^{k} a_l \right) \alpha^{k-1} \tag{49}$$

$$\frac{1}{(1-\alpha)^2} = \sum_{k=1}^{\infty} k \alpha^{k-1}. \tag{50}$$

Hence

$$(1-\alpha) \sum_{k=1}^{\infty} \alpha^{k-1} a_k - p = (1-\alpha)^2 \sum_{k=1}^{\infty} \left(\frac{1}{k} \sum_{l=1}^{k} a_l - p \right) k \alpha^{k-1}. \tag{51}$$

For any $\epsilon > 0$, choose N_ϵ such that

$$\frac{1}{N} \sum_{n=1}^{N} a_n \geqslant p - \epsilon$$

for all $N \geqslant N_\epsilon$. Then the right-hand side of (51) is bounded below by

$$\begin{aligned}
&(1-\alpha)^2 \left(\sum_{k=1}^{N_\epsilon - 1} \left(\frac{1}{k} \sum_{l=1}^{k} a_l - p \right) k\alpha^{k-1} - \epsilon \sum_{k=N_\epsilon}^{\infty} k\alpha^{k-1} \right) \\
\geqslant \ & (1-\alpha)^2 \left[\left(N_\epsilon \max_{1 \leqslant k \leqslant N_\epsilon} \left| \sum_{l=1}^{k} a_l - kp \right| \right) - \epsilon (1-\alpha)^{-2} \right] \geqslant -2\epsilon
\end{aligned}$$

for α sufficiently close to 1. This establishes (48), and (47) is obtained similarly. ∎

Acknowledgment

The authors would like to thank the referees for their comments which improved the exposition of the results of this paper.

References

[1] E. Altman, B. Gaujal, and A. Hordijk. Admission control in stochastic event graphs. Technical Report 3179,RUL-TW-97-06, INRIA and Leiden University, June 1997.

[2] E. Altman, B. Gaujal, and A. Hordijk. Balanced sequences and optimal routing. Technical Report 3180,RUL-TW-97-08, INRIA and Leiden University, January 1998.

[3] E. Altman, B. Gaujal, A. Hordijk, and G. Koole. Optimal admission, routing and service assignment control: the case of single buffer queues. In *CDC*, Tampa Bay, Fl., dec 1998. IEEE.

[4] Marcelo Bartroli and Shaler Stidham. Multimodular triangulations. Technical Report T.R. N. UNC/OR/TR-88/7, University of North Carolina at Chapell Hill, 1988.

[5] P. Glasserman and D.D. Yao. Monotone optimal control of permutable gsmps. *Mathematics of Operation Research*, 19:449–476, 1994.

[6] B. Hajek. Extremal splittings of point processes. *Mathematics of Operation Research*, 10(4):543–556, 1985.

[7] Traffic control and congestion control in b-isdn, 1995. Perth.

[8] Z. Liu and R. Righter. Optimal load balancing on distrinuted homogeneous unreliable processors. *Journal of Operations Research*, 46:563–573, 1998.

[9] A. W. Marshall and I. Olkin. *Inequalities: Theory of Majorization and its Applications*, volume 143 of *Mathematics in Science and Engineering*. Academic Press, 1979.

[10] R. T. Rockafellar. *Convex analysis*. Princeton University press, 1972.

[11] Y. G. Sinai. *Introduction to Ergodic Theory.* Princeton University Press, Princeton, N.J., 1976.

[12] R. Sznajder and J. A. Filar. Some comments on a theorem of hardy and littlewood. *Journal of Optimization Theory and Appl.*, 1992.

[13] R.R. Weber and S. Stidham. Optimal control of service rates in networks of queues. *Advances in Applied Probability*, 19:202–218, 1987.