A stochastic knapsack problem with nonlinear capacity constraint

Marcello Sanguineti

joint work with

Marco Cello, **Giorgio Gnecco**, **Mario Marchese**

Department of Communications, Computer, and System Sciences
(DIST) – University of Genoa
Via Opera Pia, 13 – 16145 Genova (Italy)

{**marcello**, **marco.cello**, **giorgio.gnecco**, **mario.marchese**}**@dist.unige.it**

Overview

- Knapsack problem
  - Classical knapsack
  - Stochastic knapsack
  - Generalized stochastic knapsack
- Nonlinear constraints and feasibility regions
- Coordinate-convex policies and corner points
- Structural properties of the policies
- A method to improve coordinate-convex policies
- An algorithm of approximate solution
- Application to Call Admission Control in telecommunication networks
- Simulation results

Classical knapsack problem

- Knapsack of capacity C
- K classes of objects
- Each object of the class $k = 1, \ldots, K$ has a size b_k and an associated reward r_k
- The objects can be placed into the knapsack as long as the sum of their sizes does not exceed the capacity C
- Problem: place the objects inside the knapsack so as to maximize the total reward

Stochastic knapsack

- Objects belonging to each class become available randomly.
- Each accepted object has a random sojourn time.

Stochastic knapsack - Applications

- Accepting and blocking calls to a circuit-switched telecommunication system which supports a variety of traffic types (e.g., voice, video, fax, etc.).
- Parallel processing where jobs require a varying number of processors as a function of their class.
- Sharing memory where tasks from different classes require different amounts of memory.
- ...

Stochastic knapsack - Literature

Various models:

- K.W. Ross and D.H.K. Tsang: The stochastic knapsack problem. *IEEE Transactions on Communications*, 37(7):740–747, 1989.
- A. J. Kleywegt and J. D. Papastavrou: The dynamic and stochastic knapsack problem with random sized items. *Operations Research*, 49(1):26–41, 2001.
- B. C. Dean, M. X. Goemans, and J. Vondrak: Approximating the stochastic knapsack problem: The benefit of adaptivity. *Mathematics of Operations Research*, 33(4):945–964, 2008.
- ...

Stochastic knapsack - Our model

Choosing a model amounts at choosing, for each class $k = 1, \ldots, K$, the inter-arrival time of the objects and the sojourn time of the accepted objects.

- The inter-arrival time is exponentially distributed with mean value $1/\lambda_k(n_k)$.
- Every accepted object has a sojourn time independent from the sojourn times of the other objects, with mean value $1/\mu_k$.

- Constraint: $\sum_{k \in K} n_k b_k \leq C$
  - n_k: number of objects of class k inside the knapsack.
- At the time of its arrival, each object is either accepted or rejected, according to a policy.
- If put into the knapsack, an object from class k generates revenue at a positive rate r_k.
- Problem: find a policy that maximizes the average revenue, by accepting or rejecting the arriving objects in dependence of the current state of the knapsack.

Remarks

- One extreme: the classical knapsack can be viewed as the limit case of our stochastic model, by setting the arrival rates for each class equal to infinity.
- The other extreme: when the arrival rates are "small", the optimal policy would consists in offering access to an object whenever sufficient volume is available (*complete sharing*).

Generalized stochastic knapsack - Our model

Linear constraint

$$\sum_{k\in K} n_k b_k \leq C$$

replaced by the nonlinear constraint

$$\sum_{k\in K} \beta_k(n_k) \leq C$$

$\beta_k(\cdot)$: nonlinear nonnegative functions

Feasibility regions

- The sets

$$\Omega_{FR} := \{(n_1, \ldots, n_K) \in \mathbb{N}_0^K : \sum_{k \in K} n_k b_k \leq C\}$$

in the linear case and

$$\Omega_{FR} := \{(n_1, \ldots, n_K) \in \mathbb{N}_0^K : \sum_{k \in K} \beta_k(n_k) \leq C$$

in the nonlinear case, are called *feasibility regions*.

- We consider the general case of nonlinearly- constrained feasibility regions.

Boundaries off the feasibility regions

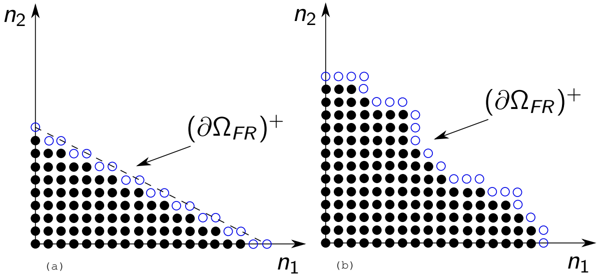


Figure: The upper boundary $(\partial\Omega_{FR})^+$ of a feasibility region Ω_{FR} with 2 classes of objects in the case of (a) a linearly-constrained Ω_{FR} and (b) a nonlinearly-constrained Ω_{FR}.

Example - Call Admission Control (CAC) as stochastic knapsack

Context: telecommunication network

- Knapsack $\leftrightarrow$ communication channel
- Object classes $\leftrightarrow$ traffic types (e.g., voice, video, etc.)
- Objects $\leftrightarrow$ requests of connections coming from different traffic types
- Objects volumes $\leftrightarrow$ bandwidth requirements
- Capacity $\leftrightarrow$ total available bandwidth
- Stochastic knapsack problem $\leftrightarrow$ optimally accepting calls in order to maximize average revenues

- Each class of users has
  - a bandwidth requirement b_k
  - a distribution for its duration
- Typically, the linear constraint $\sum_{k \in K} n_k b_k \leq C$ arises via linearizations of the nonlinear constraint $\sum_{k \in K} \beta_k(n_k) \leq C$
  - b_k: effective bandwidth of class k
- The feasibility regions model subsets of the call space $\{(n_1, \ldots, n_K) \in \mathbb{N}_0^K\}$, where given *Quality of Service* (*QoS*) constraints are satisfied

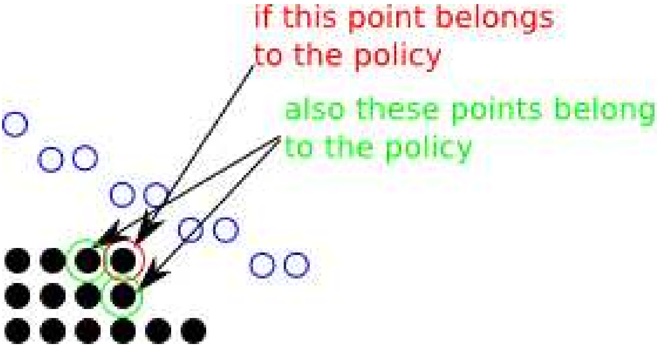
Coordinate-convex policies

- $\mathbf{e}_k$: K-dimensional vector whose k-th component is 1 and the other ones are 0
- $\mathbf{n} := (n_1, \ldots, n_k)$

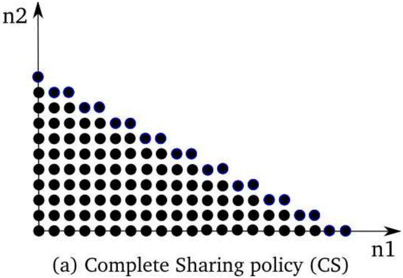
Definition 1

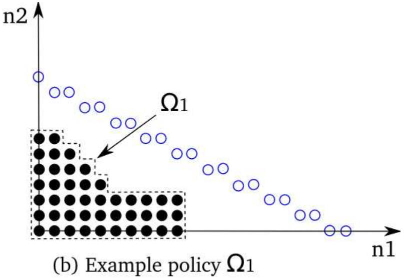
A nonempty set $\Omega \subseteq \Omega_{FR} \subset \mathbb{N}_0^K$ is a coordinate-convex set (c.c. set) iff it has the following property: for every $\mathbf{n} \in \Omega$ with $n_k > 0$ one has $\mathbf{n} - \mathbf{e}_k \in \Omega$. A policy associated with a c.c. set Ω is called a coordinate-convex policy (c.c. policy). It admits an arriving object iff after its insertion one has $\mathbf{n} \in \Omega$.

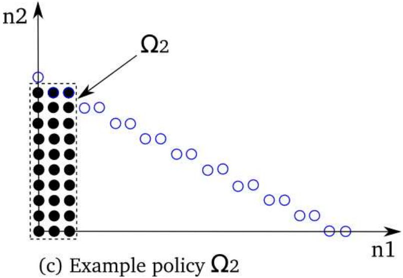
- Correspondence between c.c. sets and c.c. policies $\implies$ we use for both the symbol Ω

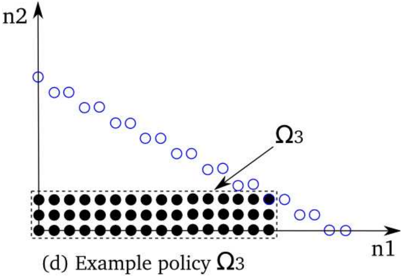

if this point belongs
to the policy
also these points belong
to the policy

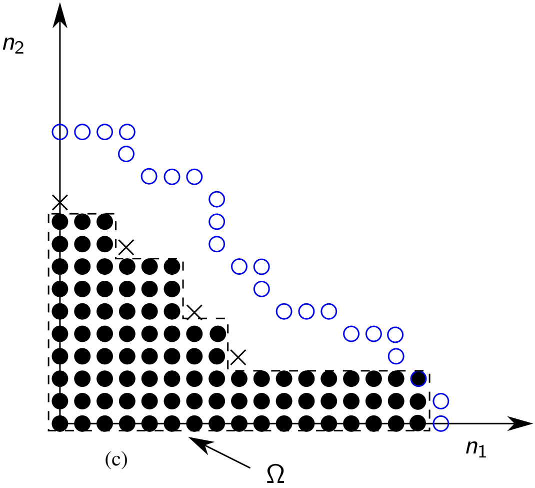
Examples of c.c. policies










n_2
n_1
(c)
Ω

Objective function

- At the time of its arrival, each object is either accepted or rejected, according to a c.c. policy
- The objective to be maximized in the set $\mathcal{P}(\Omega_{FR})$ of c.c. subsets of Ω_{FR} is given by

$$J(\Omega) := \sum_{\mathbf{n}\in\Omega} (\mathbf{n}\cdot\mathbf{r}) P_{\Omega}(\mathbf{n})$$

- $\mathbf{r} := (\mathbf{r_1}, \ldots, \mathbf{r_K})$
- $P_{\Omega}(\mathbf{n})$: steady-state probability that the current content of the knapsack is $\mathbf{n}$
- As Ω is c.c., it can be shown that $P_{\Omega}(\mathbf{n})$ takes on the product-form

$$P_{\Omega}(\mathbf{n}) = \frac{\prod_{i=1}^{K} q_i(n_i)}{\sum_{\mathbf{n}\in\Omega} \prod_{i=1}^{K} q_i(n_i)}, \text{ where } q_i(n_i) := \frac{\prod_{j=0}^{n_i-1} \lambda_i(j)}{n_i! \mu_i^{n_i}}$$

- In general, finding optimal policies is a difficult combinatorial optimization task
- Due to the form of $P_\Omega(\mathbf{n})$, in general the objective function $J(\Omega) = \sum_{\mathbf{n}\in\Omega} (\mathbf{n}\cdot\mathbf{r})P_\Omega(\mathbf{n})$ is nonlinear
- Further difficulty: given any two c.c. sets $\Omega_1, \Omega_2 \subseteq \Omega_{FR}$, in general

$$\Omega_1 \subseteq \Omega_2 \quad \not\Longrightarrow \quad J(\Omega_1) \leq J(\Omega_2)$$

Structural properties of the optimal policies

- The a-priori knowledge of structural properties of the (unknown) optimal policies
  - restricts the K-tuple $\mathbf{n} := (n_1, \ldots, n_K)$ to suitable subsets of the feasibility region Ω_{FR}
  - useful to find the solutions or, at least, good suboptimal policies
- For two classes of objects and a linear constraint, structural properties of c.c. optimal policies were derived in
  - K.W. Ross and D.H.K. Tsang: The stochastic knapsack problem. *IEEE Trans. on Communications*, 37(7):740–747, 1989

Corner points

Definition 2

The tuple $(\alpha, \beta) \in \Omega_{FR} \setminus \Omega$ *is a type-1 corner point for* Ω *iff* $\beta \geq 1$, $(\alpha, \beta - 1) \in \Omega$, *and either* $\alpha = 0$ *or* $(\alpha - 1, \beta) \in \Omega$.

Definition 3

The tuple $(\alpha, \beta) \in \Omega_{FR} \setminus \Omega$ *is a type-2 corner point for* Ω *iff* $\alpha \geq 1$, $(\alpha - 1, \beta) \in \Omega$, *and either* $\beta = 0$ *or* $(\alpha, \beta - 1) \in \Omega$.

- We use the term "corner point" to refer to either a type-1 or a type-2 corner point.
- Ω c.c. $\implies$ no two corner points can be on the same vertical or horizontal line.

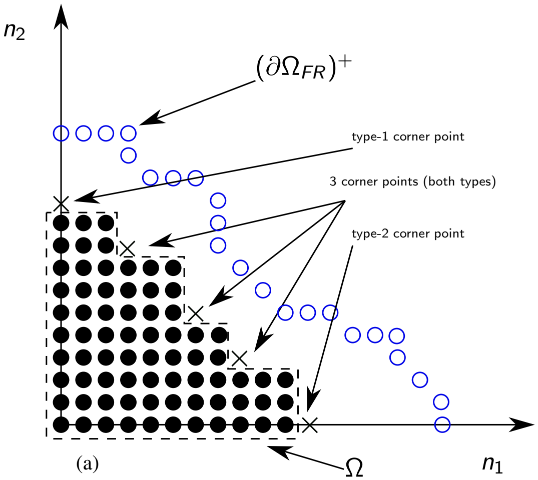


Figure: Corner points

Incremental removability and incremental admissibility

Definition 4

A nonempty set $S^- \subset \Omega_{FR}$ is incrementally removable with respect to Ω (IR_Ω) iff $S^- \subset \Omega$ and $\Omega \setminus S^-$ is a c.c. set

Definition 5

A nonempty set $S^+ \subset \Omega_{FR}$ is incrementally admissible with respect to Ω (IA_Ω) iff $S^+ \cap \Omega = \emptyset$ and $\Omega \cup S^+$ is a c.c. set.

Improving a policy

Definition 6

A c.c. policy Ω_1 *improves* Ω_2 *iff* $J(\Omega_1) > J(\Omega_2)$

- Next proposition: criterion to establish if a c.c. policy is suboptimal and, if so, to improve it

Proposition 7

Let (α, β) be a type-2 corner point for Ω and suppose that there exist $n, m, p \in \mathbb{N}_0$ such that $S^- := \{(\alpha - 1 - j, \beta + i) : j = 0, \ldots, n,\ i = 0, \ldots, p\} \subset \Omega$, is IR_Ω, and $S^+ := \{(\alpha + s, \beta + i) : s = 0, \ldots, m,\ i = 0, \ldots, p\} \subset \Omega$, is IA_Ω. Then at least one of the following inequalities holds:

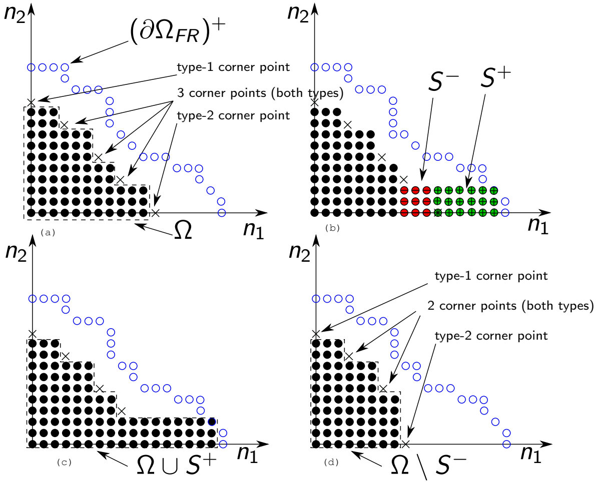
(i) $J(\Omega \cup S^+) > J(\Omega)$;

(ii) $J(\Omega \setminus S^-) > J(\Omega)$.

- Similar result for type-1 corner points.

Interpretation Proposition 7

The c.c. policy Ω is suboptimal: it can be improved by adding S^+ (Fig.(c)) or by removing S^- (Fig. (d)).



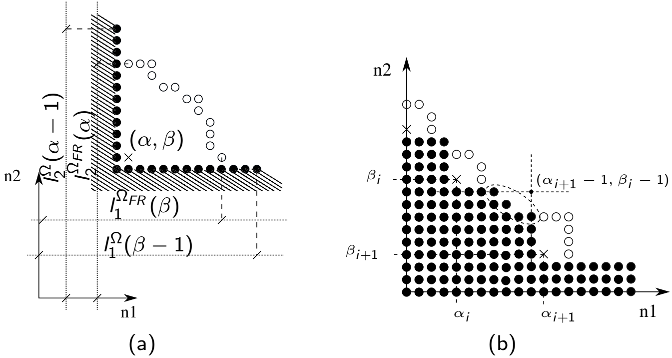


Figure: (a) An example of a corner point (α, β) for which neither Proposition 7 nor its version for type-1 corner points can be applied. (b) An example of a c.c. policy Ω^* that cannot be improved by applying Proposition 7 or its version for type-1 corner points.

- Given a type-2 corner point (α, β), it may happen that Proposition 7 can be applied for several values of the pair (m, n).
- In the next proposition, we write $S^+(m)$ and $S^-(n)$ to make the dependencies on m and n explicit.

Choosing of m and n

Proposition 8

Let the assumptions of Proposition 7 hold for $m = n = 0$ and some p.

(i) *Let $\bar{m} \geq 0$ be the maximum value of m for which $S^+(m)$ is IA_Ω. If $J(\Omega \cup S^+(0)) > J(\Omega)$, then $J(\Omega \cup S^+(m))$ is an increasing function of m for $m \in \{0, \ldots, \bar{m}\}$.*

(ii) *Let $\bar{n} \geq 0$ be the maximum value of n for which $S^-(n)$ is IR_Ω. If $J(\Omega \setminus S^-(0)) > J(\Omega)$, then $J(\Omega \setminus S^-(n))$ is an increasing function of n for $n \in \{0, \ldots, \bar{n}\}$.*

- Interpretation: the best "greedy" choices for m and n consist in taking them as large as possible: this guarantees the best local improvement of the current policy.

A greedy algorithm to improve a c.c. policy

For a c.c. policy Ω, let

$L_2(\Omega_j)$ = set of type-2 corner points for which the assumptions of Proposition 7 hold for $m = n = 0$ and some p.

- Propositions 7 and 8 suggest the following algorithm.

Algorithm 9

1. *Choose an initial c.c. policy* Ω_0*;*
2. *While* $L_2(\Omega_j) \neq \emptyset$
   let $\Omega_{j+1} \in \operatorname{argmax}\{\Omega_j \cup S^+(\bar{m}), \Omega_j \setminus S^-(\bar{n}) : (\alpha, \beta) \in L_2(\Omega_j)\}$*;*
3. *Return the current policy* Ω_j*.*

- Step 2: among the type-2 corner points of the current policy Ω_j, select the one that guarantees the largest local improvement of Ω_j.
- Such an improvement is obtained either by adding the set $S^+(\bar{m})$ to Ω_j or by removing the set $S^-(\bar{n})$.

- Algorithm 9 terminates after a finite number of iterations.
- At each step the policy that guarantees the best local improvement of the objective is selected.
- Algorithm 9 provides a c.c. policy that may not be optimal.

A property of optimal c.c. policies

- Let Ω^o denote any optimal c.c. policy (or its associated c.c. set).

Proposition 10

Ω^o has a nonempty intersection with the upper boundary $(\partial\Omega_{FR})^+$ of the feasibility region.

A property of corner points of optimal c.c. policies

- $I_2^{\Omega}(n_1) := \max\{k \in \mathbb{N}_0 \text{ such that } (n_1, k) \in \Omega\}$: maximum number of type-2 connections allowed in Ω when we have already n_1 type-1 connections.
- $I_1^{\Omega}(n_2) := \max\{h \in \mathbb{N}_0 \text{ such that } (h, n_2) \in \Omega\}$: maximum number of type-1 connections allowed in Ω when we have already n_2 type-2 connections.
- The functions $I_i^{\Omega}(\cdot)$ are nonincreasing.
- Let $n_{1,\max} := I_1^{\Omega_{FR}}(0)$ and $n_{2,\max} := I_2^{\Omega_{FR}}(0)$.

Proposition 11

(i) If (α, β) *is a type-*2 *corner point for* Ω^o*, then for some* $j = 1, \ldots, n_{2,\max}$

$$\alpha = l_1^{\Omega_{FR}}(j) + 1\,.$$

(ii) If (α, β) *is a type-*1 *corner point for* Ω^o*, then for some* $j = 1, \ldots, n_{1,\max}$

$$\beta = l_2^{\Omega_{FR}}(j) + 1\,.$$

- Proposition 11 extends to nonlinearly-constrained feasibility regions a property proved in Ross & Tsang for a linear constraint.

Interpretation of Proposition 11

- The feasibility region can be decomposed as the union of a finite number n_{rect} of disjoint rectangles with decreasing heights.
- Corner points of Ω^o: to be searched among the vertices of a grid determined by such rectangles.

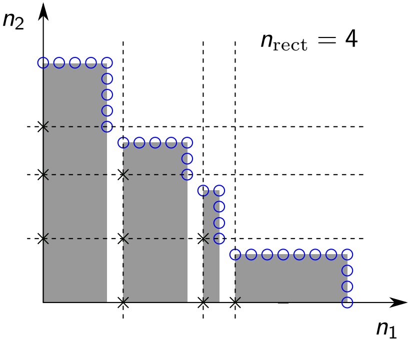


Figure: Decomposition of the feasibility region into disjoint rectangles. Crosses = potential locations of corner points of an optimal c.c. policy.

How to exploit the properties stated by Propositions 10 and 11

- Narrowing the search for the optimal c.c. policies to the ones that satisfy the necessary conditions stated in Propositions 10 and 11.
- E.g., for the feasibility region depicted in Figure 4, the next table shows how the number of candidate optimal c.c. policies drastically decreases.

All c.c. policies	> 352.716
C.c. policies satisfying Proposition 11	41
C.c. policies satisfying Propositions 10 and 11	28

Example - Call Admission Control (CAC) as stochastic knapsack

Context: telecommunication network

- Knapsack $\leftrightarrow$ communication channel
- Object classes $\leftrightarrow$ traffic types (e.g., voice, video, etc.)
- Objects $\leftrightarrow$ requests of connections coming from different traffic types (user types)
- Objects volumes $\leftrightarrow$ bandwidth requirements
- Capacity $\leftrightarrow$ total available bandwidth
- Stochastic knapsack problem $\leftrightarrow$ optimally accepting calls in order to maximize average revenues

- Each class of users has
  - a bandwidth requirement b_k
  - a distribution for its duration
- Typically, the linear constraint $\sum_{k \in K} n_k b_k \leq C$ arises via linearizations of the nonlinear constraint $\sum_{k \in K} \beta_k(n_k) \leq C$
  - $\beta_k(\cdot)$: effective bandwidth of class k
- The feasibility regions model subsets of the call space $\{(n_1, \ldots, n_K) \in \mathbb{N}_0^K\}$, where given *Quality of Service* (*QoS*) constraints are satisfied.

Simulation scenario

- Two classes of traffic: voice call traffic (class 1) and data traffic (class 2)
- Same per-call revenues of the two classes: $r_1 = r_2 = 1$.
- Class 1 and class 2 modeled by Poisson arrivals and exponential call duration
- Class-1 traffic: on average 20 calls per time unit - e.g., per minute ($\lambda_1 = 20$), with an average holding time of 3 time units per call ($\mu_1 = 1/3$).
- Class-2 traffic: $\lambda_2 = 10$ and $\mu_2 = 1/20$.

Poisson processes

- A Poisson process is a sequence of events "randomly spaced in time".
- Examples
  - Customers arriving to a bank
  - Geiger counter clicks
  - Packets arriving to a buffer
- The rate λ of a Poisson process is the average number of events per unit time (over a "long time").

- In a Poisson process, the probability of N arrivals in t units of time is given by

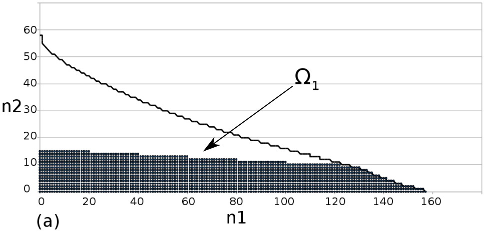
$$P_N(t) = \frac{(\lambda\, t)^N}{N!} e^{-\lambda\, t}$$

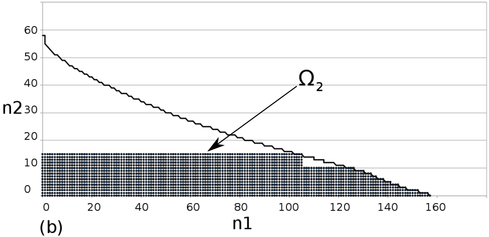
- For two intervals (s_1, s_2) and (s_3, s_4) such that $s_1 < s_2 \leq s_3 < s_4$, the number of arrivals in (s_1, s_2) is independent of the number of arrivals in (s_3, s_4).

- Feasibility region Ω_{FR} used in the simulations: has a nonlinear upper boundary that models QoS constraints. It is the one used, e.g., in
  - T. Javidi and D. Teneketzis: An approach to connection admission control in single-hop multiservice wireless networks with QoS requirements. *IEEE Transactions on Vehicular Technology*, vol. 52, pp. 1110–1124, 2003 [Figure 3].
  - M. Marchese: *QoS Over Heterogeneous Networks* Wiley, 2007 [pp. 46-49].

Simulation results

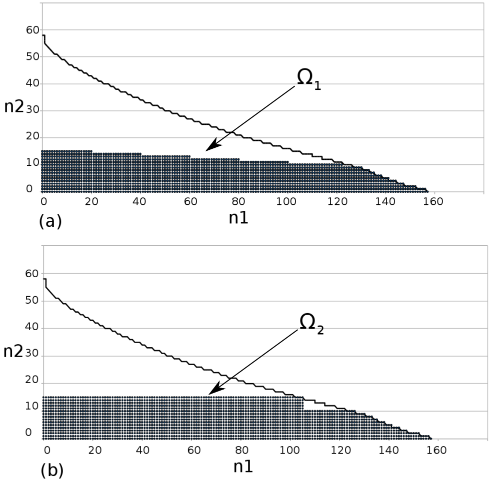
- Initial c.c. policy Ω_1.
- C.c. policy Ω_2 obtained by four iterations of Algorithm 9.





- At each iteration
  - the corner point to which Proposition 8 has been applied has been chosen according to Step 2 of Algorithm 9
  - the values of m or n have been chosen "optimally" according to Proposition 8 (i.e., $m = \bar{m}$ and $n = \bar{n}$).
- Initial value of the objective: $J(\Omega_1) = 66.4229$
- Final value: $J(\Omega_2) = 74.9198$, with an improvement of 12.8%.

- The coordinate-convex policy Ω_2 cannot be further improved via Proposition 7
- Its corner points $(0, 16)$ and $(106, 11)$ satisfy Proposition 11



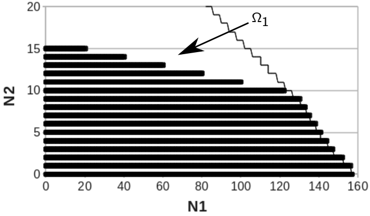


Figure: (a) initial c.c. policy Ω_1; (b) final c.c. policy Ω_2.

Conclusions

- A model of stochastic generalized knapsack
- Coordinate-convex policies
- Improving policies
- A greedy algorithm
- Application to Call Admission Control
- Simulation results for 2 classes of traffic