15th Triennial World Congress, Barcelona, Spain

AN OPERATIONAL APPROACH TO BUDGET-CONSTRAINED RELIABILITY ALLOCATION

N. Eva Wu†, Xiaoxia Wang†, Meera Sampath‡, Gregory Kott‡

†*Dept. of EE, Binghamton Univ., Binghamton, NY 13902, USA*
evawu,xwang1@binghamton.edu, Tel: (607)777-4856
‡*Xerox Wilson Center, Webster, NY 14580, USA*
msampath,gkott@crt.xerox.com, Tel: (716)422-4058

Abstract: In this paper the problem of maximal increase of system reliability is formulated as a resource allocation problem under a budget constraint. Dynamic programing is used for the optimal solution. Time to system failure is dictated by a Markov process. The system is composed of several subsystems. Each subsystem has several possible configurations that exhibit different levels of fault tolerance and incur different incremental costs at different times. Configuration dependence among subsystems is allowed. An example resembling a fault tolerant industrial process is presented, for which the proposed algorithm is used to obtain a set of maximally reliable solutions corresponding to a set of budget constraints.

Keywords: fault tolerant systems, reliability, Markov process, constrained optimization, dynamic programing

1. INTRODUCTION

Increased usage and more sophisticated features set ever demanding reliability requirements on some engineering systems, such as aircraft (Belcastro, 2001) and printing engines (Sampath, 2001). More public awareness on product and service quality, more intense business competition, and higher cost of warranty programs and liability also give industries more incentive to build more reliable engineering systems. On the other hand, development of most engineering systems is progressive, sometimes rather slowly. Therefore reliability allocation is posed in this paper as a problem of retrofitting existing systems for improved reliability rather than a problem of initial design for reliability for new systems. Correspondingly, our objective is to maximize a system reliability measure under a budget constraint rather than to minimize a cost measure in achieving a prescribed reliability goal. The principle established in this paper however, should be applicable to combinations of all cases.

In designing or modifying a complex engineering system, both reliability and life cycle cost must be considered. System reliability is the probability that the system can perform its intended mission when operated under specified conditions. Life cycle cost typically includes expenses associated with acquisition, operation, failure, etc. In general, the reliability of a complex system is a function of its component or subsystem reliabilities.

$$R(t) = f(R_1(t), \cdots, R_n(t)).$$

If a subsystem is defined as a functional unit that would cause a system level failure if it fails, then the above relation becomes

$$R(t) = \prod_{i=1}^{n} R_i(t). \tag{1}$$

Figure 1 shows two examples of such functional-based system decompositions: a flight control system and a xerographic process. It is now apparent as to how the reliability of a subsystem would affect the reliability of the overall system. A sound design should contain no redundant subsystems defined in this sense. On the other hand, each subsystem may contain many components which

are configured in various ways for fulfilling their functionalities and for providing enhanced fault tolerance. (See, for example, Thybo and Blanke, 1998.)

Reliability allocation has been posed traditionally as a least cost problem as follows (Ebeling, 1997). Suppose for each subsystem the cost-reliability relation $C_i(R_i)$ is known. The goal is to determine the most desirable subsystem reliability R_i that minimizes

$$\sum_i C_i(R_i),$$

subject to

$$\prod_i R_i \geq R^*, \;\; R_{i,min} < R_i < 1, \; i = 1, \cdots, \; n,$$

where R^* is the system reliability goal specified at some time t. The first difficulty encountered is the assumed cost-reliability relations for the subsystems, which are generally not known. However, if the cost-reliability relations can be described in a convex form such as

$$C_i(R_i) = c_i(R_i - R_{i,min})^2, i = 1, \cdots, \; n,$$

the generalized Lagrangian optimization method (Everett, 1963) can be used to find the solutions when they exist. On the other hand, even if the solutions exist, there may not be practical ways to interconnect available components into configurations that implement the optimal solutions, let alone the case of dependent configurations among subsystems. Such configuration dependence may be the result of an interface requirement, a physical space or location constraint, etc. There are some common sense approaches to reliability allocation, such as ARINC method and AGREE method (Ebeling, 1997) that do not seek for optimality. These simpler approaches, however, do not address the above issues either.

Some of the common shortfalls of the Lagrange method have been observed by researchers in the field of lossy data compression (Ortega and Ramchandran, 1998), where designers of data quantizers must trade-off between coding rate and source fidelity (rate-distortion), instead of dealing with the cost-reliability trade-offs. An operational approach to rate-distortion solution abandons the search for the best quantization solution for *any system*. It instead focuses on specific systems using specific coding schemes for which the operational rate-distortion relation can be exactly established and implemented. Perhaps the most efficient operational approach to solving rate-distortion problem is dynamic programing (Bellman, 1957). The term "operational" is adopted here in this paper for the reasons that only available configuration options for each subsystem are considered, and that dynamic programing is used for determining the optimal reliability allocation.

The paper is organized as follows. Sections 2.1 and 2.2 discuss reliability evaluation and cost assessment for a subsystem. Section 2.3 formally states the budget-constrained reliability allocation problem based the operational cost-reliability relation established, and discusses its solutions. Section 2.4 presents an algorithm based on dynamic programing. Section 3 gives a simple example that applies the algorithm. The example system resembles an industrial process in the way its reliability and cost calculations are carried out.

2. BUDGET-CONSTRAINED RELIABILITY ALLOCATION

Consider a system of n subsystems. Each subsystem is designed to carry out a specific function necessary for the normal operation of the overall system, and is therefore associated with a system level failure mode. Options in configuration within a subsystem may include variation in component types and ages, in interconnections, in levels of redundancy, and in schemes of redundancy management. These options are aimed at enhancing the fault tolerance of the subsystem. Each configuration has a cost associated with it. A set of operating base line configurations for subsystems is assumed to be already in place. The operational approach to reliability allocation also assumes that the configuration options are known to the designer, and the data needed to estimate their respective reliability and budget are available. Our goal is to select one configuration for each subsystem such that the overall system reliability is maximized under a fixed budget limit at a prespecified time. Note that the number and individual functions of subsystems are fixed in our retrofitting problem while these may alter in a new design.

2.1 *Reliability calculation*

Reliability modeling is a process of identifying the structure function of a given subsystem comprised of, say, L components with positive random lifetimes. A component or the subsystem is in state "0" (intact) before its lifetime and state "1" (failed) after its lifetime. The structure function defines a mapping: $\{0,1\}^L \rightarrow \{0,1\}$ (Aven and Jensen, 1999). Reliability assessment can be regarded as a process of evaluating the mapping, given state transition probabilities. The fundamental assumption of a Markov process is that the probability that the subsystem will undergo a transition from one state to another state depends only on the current state of the system and not

any previous states the subsystem may have experienced. Therefore, in general the subsystem has 2^L states, and 2^{2L} transition probabilities. The time the subsystem stays at a particular state is called a holding time. Holding time is a random variable. Depending on the configuration, some of the states are exit states. The sum of holding time probabilities of all exit states is the failure probability of the subsystem. Since computing such a failure probability is a mature technique and many software tools are available (Wu, 2001, and references therein), the detail of the holding time probability computation is omitted.

Let the holding time probabilities of the exit states for the jth configuration of the ith subsystem be $p_1^{(i,j)}(t), \cdots p_K^{(i,j)}(t)$. In the following development, it is assumed that holding time probabilities for all subsystems of all possible configurations have been computed. Then the reliability for the given subsystem is

$$R_{i,j}(t) = 1 - \sum_{k=1}^{K} p_k^{(i,j)}(t). \tag{2}$$

Define an alternative quantity $\lambda_{i,j}$ called an equivalent hazard rate for the subsystem

$$\lambda_{i,j}(t) \equiv -\frac{1}{t} \ln(R_{i,j}(t)). \tag{3}$$

$\lambda_{i,j}(t)$ becomes independent of time only if the subsystem has a single component configuration that has an exponential holding time distribution. Combining (1) and (3) yields the equivalent system hazard rate

$$\lambda_{j_1,\, j_2,\, \cdots,\, j_n}(t) = \sum_{i=1}^{n} \lambda_{i,j_i}(t), \tag{4}$$

where $j_i \in \{1, 2, \cdots, m_i\}$. A particular set of $(j_1,\ j_2,\ \cdots,\ j_n)$ signifies a choice of a system configuration, and hence a reliability allocation among possibly $m_1 \times m_2 \times \cdots \times m_n$ choices. (4) is now used as a performance index for reliability allocation, i.e.,

$$J_t(j_1,\ j_2,\ \cdots,\ j_n) \equiv \lambda_{j_1,\, j_2,\, \cdots,\, j_n}(t). \tag{5}$$

Once the reliabilities are allocated, the composite reliability of the system is given by (1), or with the up-to-date notations, by

$$R_{j_1,\, j_2,\, \cdots,\, j_n}(t) = \prod_{i=1}^{n} R_{i,j_i}(t). \tag{6}$$

2.2 *Cost calculation*

Suppose there are L components in the ith subsystem with the jth configuration. To simplify our discussion, only two cost items are considered. They are the initial component acquisition cost $A^{(i,j)}$ including of all components in the subsystem, and the failure cost $F_k^{(i,j)}$ which is only associated with the exit state k, and is obtained by summing up the corresponding costs of all the components which fail when exit state k is reached. Other items within the category of failure cost, such as labor, part replacement, loss of profit, etc., are not distinguished. Denote by $C^{(i,j)}(t)$ and by $C_t(j_1,\ j_2,\ \cdots,\ j_n)$ the life cycle cost of the subsystem and that of the system, respectively. Then,

$$C^{(i,j)}(t) = A^{(i,j)} + \sum_{k=1}^{K} F_k^{(i,j)} p_k^{(i,j)}(t) \tag{7}$$

$$C_t(j_1,\ j_2,\ \cdots,\ j_n) = \sum_{i=1}^{n} C^{(i,j_i)}(t), \tag{8}$$

The notion of base line configuration is now defined. Among all possible configurations with a general identifier $(j_1,\ j_2,\ \cdots,\ j_n)$, the identifier $(1,\ 1,\ \cdots,\ 1)$ is designated to the original or the base line configuration for the system. The incremental cost with respect to the baseline configuration for a subsystem is

$$\Delta C^{(i,j)}(t) = C^{(i,j)}(t) - C^{(i,1)}, \tag{9}$$

and for the overall system is

$$\Delta C_t(j_1,\ j_2,\ \cdots,\ j_n) = \sum_{i=1}^{n} (C^{(i,j_i)} - C^{(i,1)})(t), \tag{10}$$

The budget constraint can be expressed as

$$\Delta C_t(j_1,\ j_2,\ \cdots,\ j_n) \le \Delta \bar{C}, \tag{11}$$

where $\Delta \bar{C}$ is the imposed budget limit for the effort of fault tolerance enhancement. One important feature of our problem formulation is that the budget limit is fixed only in a relative sense. In terms of life-cycle cost, the budget constraint is time-varying, following the trend of the base line configuration.

2.3 *Problem statement and solutions*

Our interest is to improve the overall system reliability by using an alternative configuration that does not exceed the budget limit $\Delta \bar{C}$. Since the system reliability $R_{j_1,\, j_2,\, \cdots,\, j_n}(t)$ is in general a nonincreasing function of time, and the system life-cycle cost $C_t(j_1,\ j_2,\ \cdots,\ j_n)$ is in general a nondecreasing function of time, it is meaningful that an allocation is made for a specified time, such as the desired design life T_d of the base line system. A design life is defined to be the time

to failure that corresponds to a specific reliability, in this case, the base line system reliability $\underline{R} = R_{1,1,\cdots,1}(T_d)$. Our objective is to determine an optimal configuration identified by index set $(j_1^*,\ j_2^*,\ \cdots,\ j_n^*)$ satisfying

$$(j_1^*,\ j_2^*,\ \cdots,\ j_n^*) = \arg J_{T_d}^*. \tag{12}$$

The above implies that the optimal configuration minimizes (5) subject to (11) at T_D, i.e.,

$$\Delta C_{T_D}(j_1,\ j_2,\ \cdots,\ j_n) \le \Delta\bar{C},$$

therefore leads to the minimum equivalent hazard rate (maximum reliability) for the overall system:

$$\begin{aligned} J_{T_d}^* &= \min_{(j_1,\ j_2,\ \cdots,\ j_n)} J_{T_D}(j_1,\ j_2,\ \cdots,\ j_n) \\ &= \min_{(j_1,\ j_2,\ \cdots,\ j_n)} \sum_{i=1}^{n} \lambda_{i,j_i}(T_D). \end{aligned}$$

If a solution exists, the optimal incremental cost is given by

$$\Delta C_{T_D}(j_1^*,\ j_2^*,\ \cdots,\ j_n^*).$$

Note that a solution exists as long as $\Delta\bar{C} > 0$, or, equivalently, the total budget limit

$$\bar{C}_{T_D} = \Delta\bar{C} + \sum_{i=1}^{n} C^{(i,1)}(T_D) \tag{13}$$

is greater than the original budget allocated to the base line configuration at T_D. With the optimal configuration in place,

$$\sum_{i=1}^{n} \lambda_{i,j_i^*}(T_D) \le \sum_{i=1}^{n} \lambda_{i,1}(T_D)$$

is guaranteed, which implies, according to (3)

$$R_{j_1^*,\ j_2^*,\ \cdots,\ j_n^*}(T_D) = \prod_{i=1}^{n} R_{i,j_i^*}(T_D) \ge \prod_{i=1}^{n} R_{i,1}(T_D).$$

Therefore, the system will either have a higher reliability at the original design life, or achieves the same base line reliability $\bar{R}$ with an extended design life.

2.4 *Solution by dynamic programing*

The operational solution to the budget-constrained reliability allocation problem can be obtained in principle by enlisting all possible configurations, calculating the system reliabilities and the corresponding incremental costs for all, ordering the results in descending reliabilities, searching for the highest reliability on the list that has an acceptable budget. The corresponding configuration is the sought optimal reliability allocation.

A more systematic and efficient search method for the optimal solution is needed if the system has many functional units, and many configuration options. In this regard, dynamic programing (Bellman, 1957) becomes a natural candidate. The application of dynamic programing to reliability allocation can be explained using a trellis diagram shown in Figure 2. A trellis diagram has been used to decode convolutional codes (Forney, 1973), and to optimally trade-off rate-distortion relation in data compression (Ortega and Ramchandran, 1998).

A tree structure is first created to represent all possible solutions. Index (i, j) by each node identifies the corresponding subsystem and configuration the node is associated with. Each node of the tree at a given subsystem index represents a possible cumulative incremental expenditure. Each branch has an equivalent hazard rate corresponding to the particular configuration, and therefore as one traverses the tree from the root to the leaves, the accumulated hazard rate can be computed for each of the solutions. The Bellman's principle of optimality is applied at every subsystem index i by comparing all the accumulated hazard rates leading to the same node identifier which indicates the same subsystem configuration. Only the solution of the lowest cumulative hazard rate is retained, and the rest are removed or pruned from the tree. The principle of optimality states that the sequences of branches that result in losers so far will be the loser paths overall. This is where the computational saving is gained.

There are two issues that may complicate the solution procedure. One is caused by the budget constraint, and the other is caused by the dependence of configurations among subsystems. If the configuration of the lowest cumulative hazard rate corresponds to a node that surpasses the cumulative budget limit, the branch leading to the node must be pruned, and the next best solution in terms of the cumulative hazard rate must be retained instead. The exercise of traversing through the leaves will continue. The dependence of configurations among subsystems, on the other hand, can be dealt with by making a list of paths or a sequences of nodes that should be excluded from the search and prune the currently selected branch if it in some way matches an item on the list. Alternatively, since the trellis diagram includes all possible solutions, configuration dependence is built into the tree during its creation.

Define the minimal accumulated equivalent hazard rate for configuration j_i of the ith subsystem, and the corresponding cumulative incremental cost, repectively, as follows

$$J_i^*(j_i) = \min_{(j_1,\ j_2,\ \cdots,\ j_{i-1})} J(j_1,\ j_2,\ \cdots,\ j_i). \tag{14}$$

Variable t has been suppressed here. The optimal allocation algorithm is now stated.

Storage:
(i, j_i);
j_i^*;
$J_i^*(j_i)$, $j_i = 1, \cdots, m_i$;
ΔC_i^*;
Initialization:
$j_0^* = 1$;
$J_0(j_0^*) = 0$;
$\Delta C_0^* = 0$;
Recursion:
$i = 1 : n$;
$j_i = 1 : m_i$;
$\Delta C_i(j_i) = \Delta C_{i-1}^* + \Delta C^{(i,j_i)}$;
$\Delta C_i(j_i) > \Delta \bar{C}$, $\Rightarrow \lambda_{i,j_i} = \infty$;
$J_i^*(j_i) = J_{i-1}^*(j_{i-1}^*) + \lambda_{i,j_i}$;
$j_i = j_i + 1$;
$j_i^* = \arg\{\min_{j_i} J_i^*(j_i)\}$;
$\Delta C_i^* = \Delta C_i(j_i^*)$;
$i = i + 1$;
Output:
$(j_1^*, j_2^*, \ldots, j_n^*)$;
$J_n^*(j_n^*)$;
ΔC_n^*;

3. AN EXAMPLE

The base line system considered in this example has three single-component subsystems, each has a reliability model of exponential distribution with hazard rate $\lambda_{i,1}(t) = \lambda = 0.2$ per million hours. Alternative configuration considered is a cold spare of the same type of component in additin to the base line component for each subsystem. Coverage (Wu, 2001) for managing the redundant component in case of a component failure is 100%. Moreover, the failure rate for a component during standby is 0. The Markov models for a base line subsystem and its fault tolerant alternative are shown in Figure 3. Note that in each configuration there contains only one exit state. The failure probabilities and the equivalent hazard rates ($\lambda_{i,j}(t)$) for the two configurations of the i subsystem are shown in Figure 4.

The original acquisition costs for the base line and the fault tolerant configurations are $A^{i,1} =$ \$200 and $A^{i,2} =$ \$400, respectively. The failure costs associated with the exit state for the two configurations are $F_1^{i,1} =$ \$240 and $F_1^{i,2} =$ \$480, respectively. Substituting the hazard rate curve into (7) and (9) yield the subsystem cost curves, as shown in Figure 5.

The algorithm in the previous Section is applied and the optimal solution is sought for t between 0 to 10 million hours. Figure 6 and Figure 7 show, respectively, the incremental expenditure and the equivalent hazard rate of the system. For example, at $t = 5.025$ million hours, which is the mean time to failure (MTTF) of the base line configuration, it is found that the optimal system configuration identifiers are $(j_1^*, j_2^*, j_3^*) = (1, 1, 1)$, $(1, 2, 2)$, $(1, 1, 2)$, and $(2, 2, 2)$, under a \$10, \$500, \$1000, and \$1500 incremental budget limit, respectively. These allocations remain the same between the MTTF (approximately 5 million hours) of the base line configuration and the MTTF of the fullest redundant configuration (approximately 10 million hours).

4. ACKNOWLEDGMENT

This work was supported in part by the Xerox Corporation under Grant #HE 1321-98, in part by the NASA under Cooperative Agreement # NCC-1-336, and in part by the NSF under Grant # ECS-9615956.

The first author would like to thank Prof. Mark Fowler of Electrical Engineering at Binghamton University for introducing to her the field of data compression, which led to the idea of using an operational method for reliability allocation.

References

Aven, T., and U. Jensen, (1999). *Stochastic Models in Reliability*, p.2. Springer-Verlag.

Belcastro, C. and C. Belcastro (2001) Application of failure detection, identification, and accommodation methods for improved aircraft safety, *Proc. American Control Conference.*

Bellman, R. (1957). *Dynamic Programing,* Princeton University Press.

Ebeling, C. E. (1997). *An Introduction to Reliability and Maintainability Engineering*, pp.152-157. The McGraw-Hill Companies, Inc.

Everett, H. (1963). Generalized Lagrange multiplier method for solving problems of optimum allocation of resources, *Operations Research*, **11**, pp.399-417.

Forney, G.D. (1973). The Viterbi algorithm, *Proc. of the IEEE,* **61**. pp.268-278.

Ortega, A. and K. Ramchandran (1998). Rate-distortion methods for image and video compression, *IEEE Signal Processing Magazine,* . pp.23-50.

Sampath, M. (2001). A hybrid approach to failure diagnosis of industrial systems, *Proc. American Control Conference.*

Thybo, C. and M. Blanke (1998). Industrial Cost-Benefit Assessment for Fault-tolerant Control Systems, *Proc. IEE Conference Control'98.*

Wu, N.E. (2001). Reliability of fault tolerant control systems, *Proc. Conference on Decision and Control.*

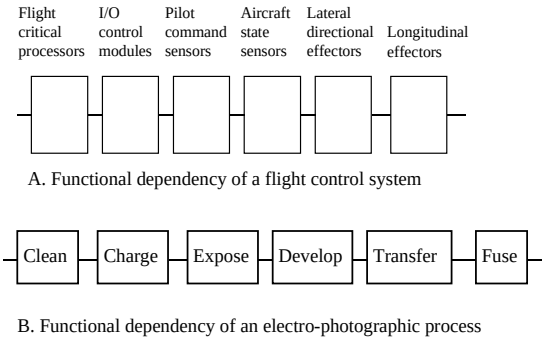

Figure 1. Examples of coarse reliability models

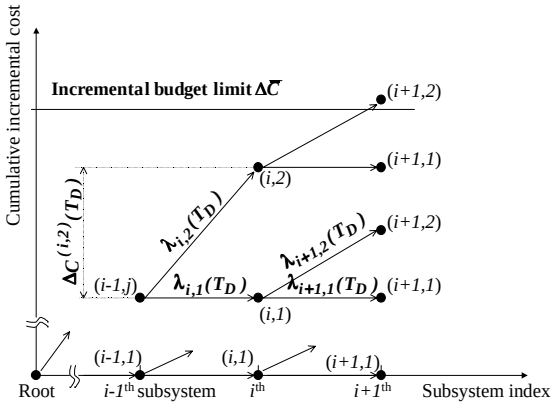

Figure 2. Trellis diagram for reliability allocation

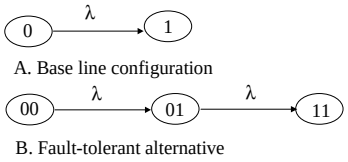

Figure 3. Rate diagrams for subsystem configurations

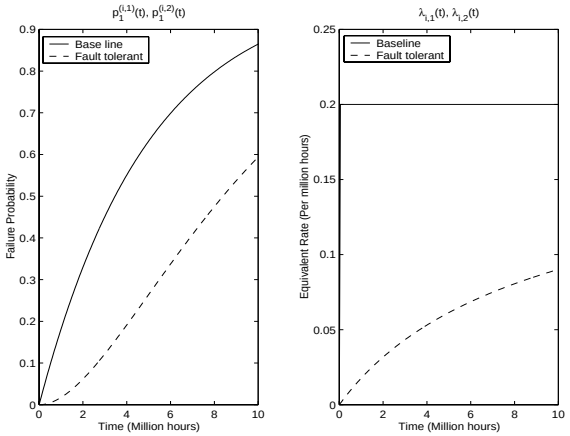

Figure 4. Subsystem failure probability and equivalent hazard rate

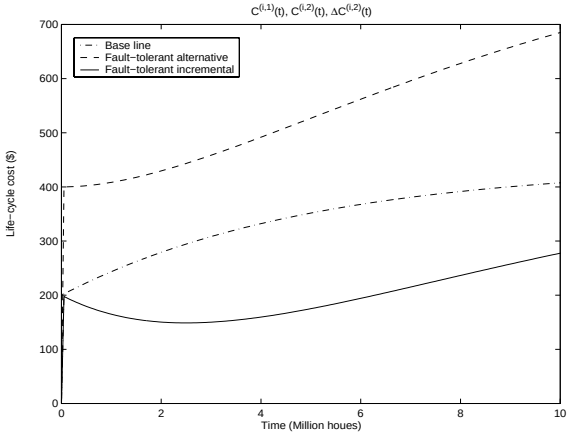

Figure 5. Subsystem life-cycle and incremental costs

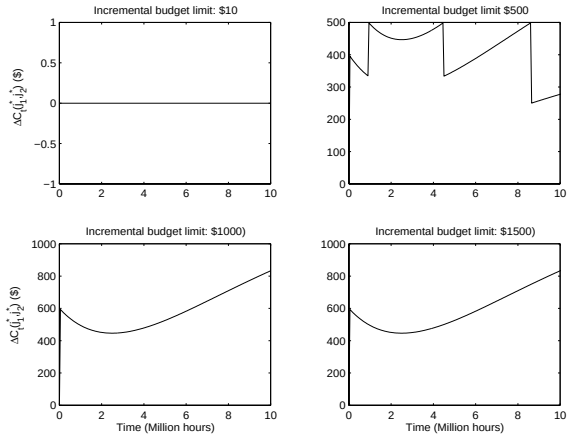

Figure 6. System incremental cost at optimal reliability allocation

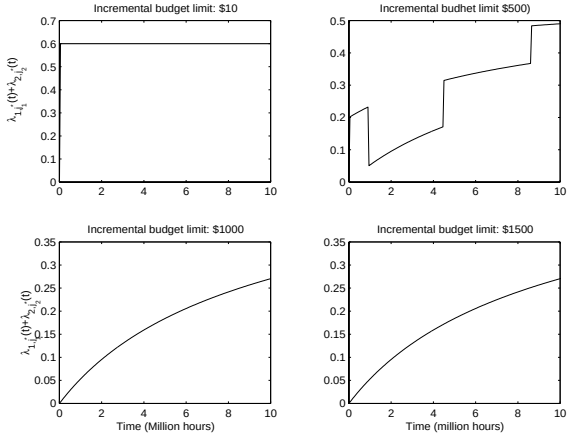

Figure 7. Minimal system equivalent hazard rate