2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

Optimality of a Hedging-Point Control Policy for a Failure-Prone Manufacturing System under a Probabilistic Cost Criterion

Amir Ahmadi-Javid and Roland Malhamé

Abstract—**Bielecki and Kumar (1988) established the optimality of a critical inventory policy (hedging policy) in a Markovian failure-prone manufacturing system subject to a constant rate of demand for parts, and for a long-term average cost structure including parts storage and demand backlog costs. Under the same conditions, and if instead of minimizing the long-term average cost, one aims at minimizing a long-term probabilistic risk measure of the running cost exceeding a given fixed barrier, we show that the optimal policy remains of the critical inventory type, albeit with different characteristics. Application of the Hamilton-Jacobi-Bellman (HJB) equation to establish optimality for this risk-averse criterion is particularly problematic. Instead, the result is established using novel, more direct arguments.**

I. INTRODUCTION

In the past three decades, the problem of finding the optimal control policy for various failure-manufacturing systems with different criteria has been studied by numerous authors. Kimemia and Gershwin [7] first developed the mathematical modeling and control theoretic framework for failure-prone multi-part manufacturing systems, and presented arguments for the plausible optimality of hedging (critical inventory) policies. For two-state systems, Akella and Kumar [2] proved the optimality of a hedging-point policy for a discounted-cost criterion in 1986; this result was subsequently extended by Bielecki and Kumar [3] for a long-term average cost criterion in 1988. After these seminal papers, the optimality of hedging-point policies was also established for other systems (see [9] and the references within for a thorough survey of this topic). A combination of optimality characteristics, intuitive appeal, and ease of implementation make hedging policies particularly popular. An examination of the literature suggests that virtually all approaches to prove the optimality of hedging-point policies hinge on a study of the associated Hamilton-Jacobi-Bellman (HJB) dynamic programming equation. As expected, success in proving optimality strongly depends on inherent "good" properties of the cost criteria (convexity, continuity, etc.).

In this paper, we consider optimization of the production laws for a failure-prone manufacturing system, under a probabilistic risk criterion. The criterion is of particular interest because of its relationship to important indicators in finance, such as value-at-risk. Recognizing that this optimality criterion lacks the properties that render feasible the standard optimality proofs based on the HJB equation feasible (see Section V for more details), we instead develop a direct and novel approach to optimality analysis tailored for the problem, and in the process extend the class of criteria for which hedging policies remain optimal. See [1, 5, 6] for other inventory problems with risk averse criteria.

The remainder of the paper is organized as follows. In Section II, we introduce the notation for our failure-prone manufacturing system, and formulate our optimal control problem. Section III develops a necessary condition for the optimality of an admissible control policy, based on which, in Section IV, we establish that the optimal policy is of the hedging type. Section V is our conclusion.

II. PROBLEM DESCRIPTION

We consider a two-mode Markovian machine with one operating mode, $\alpha = 1$, and one failure mode, $\alpha = 0$. By $\alpha(t)$ and $X(t)$, we respectively denote the mode and the surplus (inventory if positive, and backlog if negative) processes associated with the machine at time t. The mode $\alpha(t)$ is assumed to evolve according to a continuous-time Markov chain, where q_1 is the rate of transition from operating to failure mode, and q_0 is the rate of transition from failure to operating mode. The demand rate is assumed to be a constant d and backlog is allowed. We limit our analysis to *stationary* state-feedback production policies. More specifically, the surplus process $X(t)$ satisfies the equation

$$\frac{\mathrm{d}X(t)}{\mathrm{d}t} = u(X(t), \alpha(t)) - d \tag{1}$$

where $u : \Re \times \{0,1\} \to \Re^+$, and the set of admissible feedback policies is defined by

$$u(x,i) = \begin{cases} r_+(x) & x \ge 0,\ i = 1 \\ r_-(x) & x < 0,\ i = 1 \\ 0 & x \in \Re,\ i = 0 \end{cases} \tag{2}$$

with $0 \le u(x,1) \le k$ for all $x \in \Re$, and where $r_+(x)$, $r_-(x)$ are arbitrary Borel measurable functions as long as they induce ergodicity in the controlled hybrid-state Markov process $(X(t), \alpha(t))^T$ (see [8] for more details).

The objective is to determine an admissible control policy as defined in (2), so as to minimize the following long-run expected average cost (assumed to exist under the

Amir Ahmadi-Javid is with Department of Industrial Engineering, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran.

Roland Malhamé is with GERAD and Department of Electrical Engineering, Ecole Polytechnique de Montréal, Montreal, Canada.

ergodicity of the controlled process)

$$J_r = \lim_{T\to+\infty} \mathrm{E}\left\{\frac{1}{T}\int_0^T I_{(l,+\infty)}\left(c^+X_r^+(t)+c^-X_r^-(t)\right)\mathrm{d}t\right\} \tag{3}$$

where $X_r^+(t)=\max\{X_r(t),0\}$, $X_r^-(t)=\max\{-X_r(t),0\}$, the index r refers to the control policy, and $I_{(l,+\infty)}:\Re_+\to\{0,1\}$ is the indicator function

$$I_{(l,+\infty)}(x)=\begin{cases}1 & x>l\\ 0 & x\le l\end{cases}$$

for a given positive constant l. Constants c^+ and c^- are respectively the instantaneous holding and backlog costs, and therefore, $c^+X_r^+(t)+c^-X_r^-(t)$ denotes the cost incurred per unit time at time t.

Criterion (3) can be interpreted as the long-term probability that under a given admissible production policy, the instantaneous cost will exceed a given threshold l. This threshold may represent the maximum instantaneous cost that the manufacturer is willing to tolerate for any length of time. Therefore, he/she would like to limit the probability of that level being exceeded.

A proof of the optimality of a hedging-point policy for criterion (3) is difficult to achieve using standard dynamic programming verification theorems (see Section V for more details). This is in view of the non-convexity and non-continuity of the integrand of (3). We develop more direct arguments at the cost of limiting at the outset the set of admissible policies to the set of stationary state-feedback, ergodicity-inducing policies.

Given the scalar nature of the surplus $X(t)$ and the structure of criterion (3), it turns out that, without loss of optimality, one can further specialize the structure of the set of relevant control policies in (2). Indeed, define

$$z=\sup\{y\in\Re: r(x)>d\ \forall x\le y\}$$

where

$$r(x)=u(x,1)=\begin{cases}r_+(x) & x\ge 0\\ r_-(x) & x<0\end{cases}$$

and z can possibly be infinite. Note that any finite $x_0>z$ would be transient. Indeed, with probability one, following successive machine failures, $X(t)$ will become smaller than z. In addition, by construction, $\forall\varepsilon>0$, $\exists y\in(z,z+\varepsilon)$, such that $r(x)\le d$. As a result, by continuity of its trajectories, $X(t)$ will never again reach region $x\ge x_0$. Consequently, without loss of optimality, we shall assume the following structure of on-mode feedback policies:

$$r(x)=\begin{cases}0 & x>z\\ \beta & x=z\\ f(x) & x<z\end{cases} \tag{4}$$

where z is a constant, possibly infinite, β is some constant in the interval $[d,k]$, and $d<f(x)\le k$ for all $x<z$. In addition, for simplicity, we can impose $\beta=d$, since the region above z is transient anyway. Thus, we look for optimality over the following restricted set of stationary, ergodicity-inducing state-feedback laws:

$$r(x)=\begin{cases}0 & x>z\\ d & x=z\\ f(x) & x<z.\end{cases} \tag{5}$$

III. Necessary conditions for optimality

By the ergodicity requirement on the admissible control policies

$$J_r=\mathrm{E}\left(I_{(l,+\infty)}\left(c^+Y_r^+ + c^-Y_r^-\right)\right)=\Pr\left(c^+Y_r^+ + c^-Y_r^- > l\right) \tag{6}$$

where Y_r, Y_r^+, Y_r^- are the random variables associated with the unique steady-state distributions of $X_r(t)$, $X_r^+(t)$, $X_r^-(t)$, respectively, under the production policy r.

We would like to establish that $z\ge 0$ is a necessary condition for optimality, and, furthermore, that an optimal policy must satisfy $r_-(x)=k$ a.e. for $x<0$.

One can write:

$$\begin{aligned}J_r&=\Pr\left(c^+Y_r^+ + c^-Y_r^- > l\right)\\ &=p_r^+\Pr\left(c^+Y_r^+>l\,\middle|\,Y_r\ge 0\right)+p_r^-\Pr\left(c^-Y_r^->l\,\middle|\,Y_r<0\right)\end{aligned} \tag{7}$$

where $p_r^+=\Pr(Y_r\ge 0)$ and $p_r^-=\Pr(Y_r<0)$ are the steady-state probabilities of the surplus process being respectively non-negative or negative. We have the following lemma.

Lemma 1. A necessary condition for the optimality of a feedback policy r given in (5) is that

$$\Pr\left(c^+Y_r^+>l\,\middle|\,Y_r\ge 0\right)<\Pr\left(c^-Y_r^->l\,\middle|\,Y_r<0\right). \tag{8}$$

Proof. First note that if z in (5) is negative for the corresponding policy, the result holds automatically. Thus, we assume that $z\ge 0$, and therefore $x=0$ must be positive recurrent, because, otherwise, one could assume $z<0$ with no effect on the long-term probability J_r. Now suppose that r is optimal and the condition of the lemma is not satisfied, i.e.,

$$\Pr\left(c^+Y_r^+>l\,\middle|\,Y_r\ge 0\right)\ge\Pr\left(c^-Y_r^->l\,\middle|\,Y_r<0\right). \tag{9}$$

Then, it would be possible to consider the new control policy $\tilde r$ by modifying f in the control policy (5) as follows:

$$\tilde f(x)=\begin{cases}d & x=0\\ 0 & 0<x\le z.\end{cases}$$

Note that if r is ergodicity inducing, so is $\tilde r$, and

- i This change does not affect the left term, namely $\Pr\left(c^-Y_{\tilde r}^->l\,\middle|\,Y_{\tilde r}<0\right)=\Pr\left(c^-Y_r^->l\,\middle|\,Y_r<0\right)$, because of the alternating renewal nature of non-negative and negative cycles of the surplus process;
- ii While this change decreases p_r^+ to $p_{\tilde r}^+$, $p_{\tilde r}^+$ will still remain strictly positive because we assumed positive recurrence of $x=0$;
- iii $\tilde r$ yields $\Pr\left(c^+Y_{\tilde r}^+>l\,\middle|\,Y_{\tilde r}\ge 0\right)=0$.

As a result, by (7) and our assumption (9),

$$\begin{aligned}&\Pr\left(c^+Y_{\tilde r}^+ + c^-Y_{\tilde r}^- > l\right)\\&= \left(1-p_{\tilde r}^+\right)\Pr\left(c^-Y_{\tilde r}^- > l \middle| Y_{\tilde r} < 0\right) = \left(1-p_{\tilde r}^+\right)\Pr\left(c^-Y_r^- > l \middle| Y_r < 0\right)\\&< \Pr\left(c^-Y_r^- > l \middle| Y_r < 0\right) < \Pr\left(c^-Y_r^- > l \middle| Y_r < 0\right) +\\&\quad p_r^+\left[\Pr\left(c^+Y_r^+ > l \middle| Y_r \ge 0\right) - \Pr\left(c^-Y_r^- > l \middle| Y_r < 0\right)\right]\\&= \Pr\left(c^+Y_r^+ + c^-Y_r^- > l\right),\end{aligned}$$

which is a contradiction since r was assumed to be optimal. This completes the proof. □

Theorem 1. An optimal policy r must be such that $r(x) = k$ a.e., for $x < 0$.

Proof. Assume that, for an optimal feedback policy r with a structure as in (5), there exists an interval of non-zero length $I = (a,b]$ such that $a < b < \min\{0,z\}$ and $d < r(x) < k$ for $x \in I$. Because the controlled process must be ergodic, interval I will be positive recurrent.

Let us first consider $z \ge 0$, which implies that p_r^+ is strictly positive. Then, modify the control policy on interval I so that $r(x) = k$ on I while leaving it unchanged everywhere else. Let $\tilde r$ be the corresponding feedback policy. Clearly, on every machine sample paths $X_{\tilde r}(t)[\omega]$ and $X_r(t)[\omega]$, for any ω, starting from identical initial conditions, we have $X_r(t)[\omega] \ge X_{\tilde r}(t)[\omega]$ for all $t \ge 0$, which implies $X_r(t) \ge X_{\tilde r}(t)$. Furthermore, interval I is visited infinitely often by both processes $X_r(t)$ and $X_{\tilde r}(t)$, and the sojourn time on I is finite for both, because I has a non-zero length and processes move at finite speed; so, on average, $X_{\tilde r}(t)$ will gain a definite advantage over $X_r(t)$; thus, the mean return time from $x = 0_-$ to point $x = 0$ will be strictly shorter under control policy $\tilde r$, than it is under r. Given that the mean return times from $x = 0_+$ to $x = 0$ are the same for $\tilde r$ and r, as well as for the mean first-passage times from $x = 0$, to either $x = 0_+$ or $x = 0_-$, one can conclude that $p_{\tilde r}^- < p_r^-$.

Moreover, we have

$$\Pr\left(c^-Y_{\tilde r}^- > l \middle| Y_{\tilde r} < 0\right) < \Pr\left(c^-Y_r^- > l \middle| Y_r < 0\right) \tag{10}$$

$$\Pr\left(c^+Y_{\tilde r}^+ > l \middle| Y_{\tilde r} \ge 0\right) = \Pr\left(c^+Y_r^+ > l \middle| Y_r \ge 0\right). \tag{11}$$

From the optimality of feedback policy r and (11) one should have

$$\begin{aligned}J_r &= \Pr\left(c^+Y_r^+ > l \middle| Y_r \ge 0\right)\\&\quad + p_r^-\left[\Pr\left(c^-Y_r^- > l \middle| Y_r < 0\right) - \Pr\left(c^+Y_r^+ > l \middle| Y_r \ge 0\right)\right]\\&\le J_{\tilde r} = \Pr\left(c^+Y_{\tilde r}^+ > l \middle| Y_{\tilde r} \ge 0\right)\\&\quad + p_{\tilde r}^-\left[\Pr\left(c^-Y_{\tilde r}^- > l \middle| Y_{\tilde r} < 0\right) - \Pr\left(c^+Y_{\tilde r}^+ > l \middle| Y_{\tilde r} \ge 0\right)\right]\\&= \Pr\left(c^+Y_r^+ > l \middle| Y_r \ge 0\right)\\&\quad + p_{\tilde r}^-\left[\Pr\left(c^-Y_{\tilde r}^- > l \middle| Y_{\tilde r} < 0\right) - \Pr\left(c^+Y_r^+ > l \middle| Y_r \ge 0\right)\right].\end{aligned} \tag{12}$$

We now argue that given Lemma 1, (9) and (10), inequality (12) cannot hold. Indeed two cases can possibly arise:

i $\Pr\left(c^-Y_{\tilde r}^- > l \middle| Y_{\tilde r} < 0\right) > \Pr\left(c^+Y_r^+ > l \middle| Y_r \ge 0\right)$: in this case, the right-hand side of (12) is strictly less than the left-hand side. Indeed, by Lemma 1, $p_{\tilde r}^- < p_r^-$. and (10), contradicting (12), we have

$$\begin{aligned}&p_{\tilde r}^-\left[\Pr\left(c^-Y_{\tilde r}^- \ge l \middle| Y_{\tilde r} < 0\right) - \Pr\left(c^+Y_r^+ > l \middle| Y_r \ge 0\right)\right]\\&< p_r^-\left[\Pr\left(c^-Y_r^- > l \middle| Y_r < 0\right) - \Pr\left(c^+Y_r^+ > l \middle| Y_r \ge 0\right)\right];\end{aligned}$$

ii $\Pr\left(c^-Y_{\tilde r}^- > l \middle| Y_{\tilde r} < 0\right) \le \Pr\left(c^+Y_r^+ > l \middle| Y_r \ge 0\right)$: in this case, the right-hand side of (12) is also strictly less than the left-hand side, since in the right-hand side

$$p_{\tilde r}^-\left[\Pr\left(c^-Y_{\tilde r}^- > l \middle| Y_{\tilde r} < 0\right) - \Pr\left(c^+Y_r^+ > l \middle| Y_r \ge 0\right)\right] \le 0$$

and by Lemma 1 in the left-hand side

$$p_r^-\left[\Pr\left(c^-Y_r^- > l \middle| Y_r < 0\right) - \Pr\left(c^+Y_r^+ > l \middle| Y_r \ge 0\right)\right] > 0.$$

Thus (12) cannot hold and therefore r cannot be optimal.

The case $z < 0$ can be disposed of in a similar manner. In this case, $p_r^+ = 0$ and $J_r = \Pr\left(c^-Y_r^- > l \middle| Y_r < 0\right)$. By modifying the control policy so that on both interval I and $z \le x < 0$, $r(t) = k$, setting $r(0) = d$, and leaving $r(x) = 0$ for $x > 0$, the resulting control $\tilde r$ performs better than r, so again we have

$$\Pr\left(c^-Y_{\tilde r}^- > l \middle| Y_{\tilde r} < 0\right) < \Pr\left(c^-Y_r^- > l \middle| Y_r < 0\right) \text{and } p_{\tilde r}^- < p_r^- = 1.$$

The above inequalities yield

$$J_r = \Pr\left(c^-Y_r^- > l \middle| Y_r < 0\right) > p_{\tilde r}^- \Pr\left(c^-Y_{\tilde r}^- > l \middle| Y_{\tilde r} < 0\right) + p_{\tilde r}^+ \times 0 = J_{\tilde r},$$

which again contradicts the optimality of r. This completes the proof. □

The proof presented for Theorem 1 also implies that for a policy to be optimal, one must always have $z \ge 0$. This leads us to the following corollary.

Corollary 1. For an optimal policy, it is necessary that $z \ge 0$ and, in characterization (5) of the policy, $r(x) = k$ a.e., for $x < 0$.

IV. Optimality of Hedging-Point Policy

From Corollary 1, an optimal policy within the restricted class characterized by (5) must have the structure

$$r(x) = \begin{cases} 0 & x > z \\ d & x = z \\ g(x) & 0 < x < z \\ k & x < 0 \end{cases} \tag{13}$$

for some $d < g(x) \le k$ and $z \ge 0$. In this section, we will show that $g(x) = k$.

Actually, each member r of the class of policies (13) can be represented alternatively as

$$r(x)=R_z(x)=\begin{cases}0 & x>z\\ d & x=z\\ R(x) & x<z\end{cases} \tag{14}$$

for some nonnegative number z and a non-stop production policy R introduced for convenience here, such that

$$R(x)=\begin{cases}h(x) & 0<x\\ k & x<0\end{cases} \tag{15}$$

with $d<h(x)\le k$ for all $x>0$.

Due to the equivalence between the two classes of policies (13) and (14), we will alternatively work with structure (14) in Lemma 4, Theorem 3 and Lemma 5 below.

In the sequel, assume $z\ge 0$ and $r(x)=k$ for $x<0$. Also, define function

$$\lambda(w)=\frac{q_0}{d}-\frac{q_1}{r(w)-d}.$$

To carry our analysis further, we shall first determine the steady-state distribution of the surplus level for the policy given in (13).

Theorem 2. Consider a control policy characterized by (13) with $z\ge 0$ assumed finite. The following holds:

i. The surplus process will have a steady-state distribution if and only if

$$\frac{q_0}{d}-\frac{q_1}{k-d}>0;$$

ii. If condition i holds, then the cumulative distribution of Y_r, the random variable associated with the steady-state distributions of $X(t)$, is given by

$$F_{Y_r}(y)=\begin{cases}\int_{-\infty}^{y} f_{Y_r}(x)dx & y<z\\ 1 & y\ge z\end{cases}$$

where for $x<z$

$$f_{Y_r}(x)=p(z)q_1\left(\frac{1}{d}+\frac{1}{r(x)-d}\right)\exp\left\{-\int_x^z\lambda(w)\mathrm{d}w\right\},$$

and

$$p(z)=\frac{1}{1+q_1\int_{-\infty}^{z}\left[\left(\frac{1}{d}+\frac{1}{r(x)-d}\right)\exp\left\{-\int_x^z\lambda(w)\mathrm{d}w\right\}\right]\mathrm{d}x}.$$

Proof. For part i, note that a necessary and sufficient condition for the existence of a steady-state distribution is that the mean return time of the surplus process from 0_- to 0 be finite because, due to the assumed finiteness of z, the mean return time from 0_+ to 0 will clearly be finite. Following [4], a necessary and sufficient condition for this to happen is condition i. Another proof for this condition can be obtained from the boundary conditions studied in the proof of part ii.

As for part ii, let $f_{Y_r}^{1,-}$ and $f_{Y_r}^{0,-}$ be the hybrid on-mode and off-mode steady-state probability densities of the surplus process on region $(-\infty,0)$ and $f_{Y_r}^{1,+}$ and $f_{Y_r}^{0,+}$ on region $(0,z)$ and $f_{Y_r}(x)$ the unconditional probability density on $(-\infty,z)$. Also, let $p(z)$ be the steady-state probability mass at point z. Indeed, we have

$$f_{Y_r}(x)=\begin{cases}f_{Y_r}^{1,-}(x)+f_{Y_r}^{0,-}(x) & x<0\\ f_{Y_r}^{1,+}(x)+f_{Y_r}^{0,+}(x) & 0<x<z\end{cases}$$

and $F_{Y_r}(y)=\int_{-\infty}^{y} f_{Y_r}(x)\mathrm{d}x+p(z_+)$. Note that for any Borel measurable set $A\subseteq(-\infty,z)$ we have

$$\int_A\left[f_{Y_r}^{0,-}(x)+f_{Y_r}^{0,+}(x)\right]\mathrm{d}x=\lim_{t\to\infty}\Pr\{X(t)\in A,\alpha(t)=0\}$$

$$\int_A\left[f_{Y_r}^{1,-}(x)+f_{Y_r}^{1,+}(x)\right]\mathrm{d}x=\lim_{t\to\infty}\Pr\{X(t)\in A,\alpha(t)=1\}$$

$$p(z)=\lim_{t\to\infty}\Pr\{X(t)=z\}.$$

Now we consider two cases:

Case A: The function $h(x)$ in (15) is differentiable on $(0,z)$.

By writing the steady-state forward Kolmogorov equations on $(-\infty,0)$ and $(0,z)$ we obtain

for $x<0$

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[(k-d)f_{Y_r}^{1,-}(x)\right]=-q_1f_{Y_r}^{1,-}(x)+q_0f_{Y_r}^{0,-}(x)$$

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[d\,f_{Y_r}^{1,-}(x)\right]=-q_1f_{Y_r}^{1,-}(x)+q_0f_{Y_r}^{0,-}(x)$$

for $0<x<z$

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[(h(x)-d)f_{Y_r}^{1,+}(x)\right]=-q_1f_{Y_r}^{1,+}(x)+q_0f_{Y_r}^{0,+}(x)$$

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[d\,f_{Y_r}^{1,+}(x)\right]=-q_1f_{Y_r}^{1,+}(x)+q_0f_{Y_r}^{0,+}(x).$$

Considering an additional differential equation for probability mass $p(z)$, boundary conditions at 0, z and $-\infty$, and the normalization equation we also have

$$q_1\,p(z)=(r(z_-)-d)f_{Y_r}^{1,+}(z_-)=d\,f_{Y_r}^{0,+}(z_-)$$

$$(k-d)f_{Y_r}^{1,-}(0_-)=(h(0_+)-d)f_{Y_r}^{1,+}(0_+)$$

$$-d\,f_{Y_r}^{0,-}(0_-)=-d\,f_{Y_r}^{0,+}(0_+)$$

$$\lim_{x\to-\infty}f_{Y_r}^{1,-}(x)=\lim_{x\to-\infty}f_{Y_r}^{0,-}(x)=0$$

$$\int_{-\infty}^{0}\left[f_{Y_r}^{1,-}(x)+f_{Y_r}^{0,-}(x)\right]\mathrm{d}x+\int_0^z\left[f_{Y_r}^{1,+}(x)+f_{Y_r}^{0,+}(x)\right]\mathrm{d}x+p(z)=1.$$

One can complete the calculation to establish the result. We omit the details to save space.

Case B: The function $h(x)$ in (15) is not necessarily differentiable.

Since r is bounded and admissible, it is square-integrable. Thus, by considering Carleson's theorem we can construct a sequence of smooth bounded functions $\{f_n\}_{n=1}^{\infty}$ based on the partial sum of the Fourier series that converges almost everywhere to h. Each f_n may be modified as follows:

$$h_n(x)=\frac{\sup_{0\le x<z}\{h(x)\}-\inf_{0\le x<z}\{h(x)\}-1/n}{\sup_{0\le x<z}\{f_n(x)\}-\inf_{0\le x<z}\{f_n(x)\}}\times\left(f_n(x)-\inf_{0\le x<z}\{h(x)\}\right)+\inf_{0\le x<z}\{h(x)\}+1/n.$$

By this modification, we always have $d<h_n(x)\le k$, $n=1,2,\ldots$, and again, the new sequence $\{h_n\}_{n=1}^{\infty}$ converges almost everywhere to h. Let $\{r_n\}_{n=1}^{\infty}$ be the sequence of the corresponding control policies that converges almost everywhere to r. Also, define $\{X_n(t)\}_{n=1}^{\infty}$ as the sequence of surplus processes associated with $\{r_n\}_{n=1}^{\infty}$. Suppose $Z(t)$ to be the limit of $\{X_n(t)\}_{n=1}^{\infty}$, then for all ω, by Lebesgue's dominated convergence theorem, we have (note $0\le\alpha(s)r_n\{X_n(s)[\omega]\}\le k$ for all $s\ge 0$)

$$Z(t)[\omega]=\lim_{n\to\infty}X_n(t)[\omega]=X(0)+\lim_{n\to\infty}\int_0^t[\alpha(s)r_n\{X_n(s)[\omega]\}-d]\mathrm{d}s$$
$$=X(0)+\int_0^t[\alpha(s)r\{Z(s)[\omega]\}-d]\mathrm{d}s,$$

which means that $Z(t)$ is actually the surplus process associated with r, i.e., $X_r(t)$. Thus the steady-state distribution associated with $\{X_n(t)\}_{n=1}^{\infty}$, which can be obtained from Case A, converges to that of $X_r(t)$. This completes the proof. □

We also present the following three lemmas, which correspond to long but straightforward calculations omitted here for reasons of space.

Henceforth we assume the condition given in Theorem 2.i.

Lemma 2. The function $f_{Y_r}(x)$ characterized in Theorem 2 can be represented as follows:

$$f_{Y_r}(x)=q_r(z)q_1\left(\frac{1}{d}+\frac{1}{r(x)-d}\right)\exp\left(\int_0^x\lambda(w)\mathrm{d}w\right),$$

where

$$q_r(z)=1\Bigg/\left\{1+\frac{q_0+q_1}{d}\int_0^z\left(\exp\int_0^x\lambda(w)\mathrm{d}w\right)\mathrm{d}x+q_1\int_{-\infty}^0\left[\left(\frac{1}{d}+\frac{1}{k-d}\right)\exp\left\{-\left(\frac{q_0}{d}-\frac{q_1}{k-d}\right)x\right\}\right]\mathrm{d}x\right\}.$$

Lemma 3. By defining $G_r(x)=F_{Y_r}(x)/q_r(z)$ for $x<z$, we have $J_r=1-F_{c^+Y_r^++c^-Y_r^-}(l)$ where

$$F_{c^+Y_r^++c^-Y_r^-}(v)=\begin{cases}q_r(z)\left[G_r\left(\frac{v}{c^+}\right)-G_r\left(-\frac{v}{c^-}\right)\right] & 0\le v<c^+z\\ 1-q_r(z)G_r\left(-\frac{v}{c^-}\right) & v\ge c^+z\,.\end{cases}$$

Lemma 4. If $r(x)=R_z(x)$, for a fixed number $z\ge 0$, the function $q_{R_z}(z)$ introduced in Lemma 2 is a strictly decreasing function of R, in the sense that, for two policies R and $\widetilde{R}$, if $R(x)\le\widetilde{R}(x)$, and $R(x)<\widetilde{R}(x)$ a.e., then $q_{R_z}(z)>q_{\widetilde{R}_z}(z)$. Moreover, for a fixed policy R, $q_{R_z}(z)$ is strictly decreasing in z.

The next theorem presents an interesting result that shows that non-stop production policies (15) cannot be optimal in any situation.

Theorem 3. For a fixed function R, the production policy R_z defined in (14) is optimal for $z=l/c^+$.

Proof. From Lemma 3, J_r can be rewritten as a function of $z\ge 0$:

$$J_{R_z}(z)=\begin{cases}q_{R_z}(z)G\left(-\frac{l}{c^-}\right) & 0\le z_+\le\frac{l}{c^+}\\ 1-q_{R_z}(z)\left[G\left(\frac{l}{c^+}\right)-G\left(-\frac{l}{c^-}\right)\right] & \frac{l}{c^+}<z.\end{cases}\tag{16}$$

By Lemma 4, $q_{R_z}(z)$ is a strictly decreasing function in z; thus, $J_{R_z}(z)$ is strictly decreasing on $0\le z\le l/c^+$ and strictly increasing function on $l/c^+<z$. Therefore, $z=l/c^+$ is the unique optimum. □

Lemma 5. An optimal policy in the class of polices (14), is $r(x)=R_z(x)$ with $R(x)=k$ a.e., and $z=l/c^+$.

Proof. From Theorem 3, we know that, for an arbitrary given function $h(x)$ in (15), optimal performance is attained for $z=l/c^+$. Therefore, an optimal choice of $h(x)$ is one that minimizes the performance index

$$J_{R_z}(l/c^+)=q_{R_z}(l/c^+)G(-l/c^-),$$

obtained from (16). By Lemma 4, $q_{R_z}(z)$ is a strictly decreasing function of R; thus, for the optimal policy, we have $R(x)=k$, a.e.. Note that $G(-l/c^-)$ is independent of the function $h(x)$ since $-l/c^-<0$. □

We are now ready to state our main result.

Theorem 4. **(*Optimality of a hedging policy*).** In the class of policies (5), the hedging-point policy

$$r(x)=\begin{cases}0 & x>z\\ d & x=z\\ k & x<z\end{cases}\tag{17}$$

with $z=l/c^+$ is optimal, and is also unique, in the sense that all other optimal policies with structure (5) are almost everywhere identical to it.

Proof. Note that by [2, Theorem 6] the hedging-point policy (17) is stationary, associated with a well-defined solution for the corresponding differential equation (1). In addition, under the assumption in Theorem 2.i, policy (17) is ergodicity inducing, so it is admissible. Also, recall that the two classes of policies (13) and (14) are identical. Now in view of Lemma 5 and Corollary 1, one can conclude that, in the class of policies (5), all optimal policies are almost everywhere identical to policy (17). □

V. Concluding Remarks

This paper develops the structure of an optimal policy for a failure-prone manufacturing system responding to constant demand under a probabilistic cost criterion given in (3). Admissibility of production policies is restricted at the outset to the set of stationary state-feedback policies that produce a well-defined solution for the controlled surplus process, and induce its ergodicity. It is shown that, for a restricted class of feedback policies, including the optimal policy, long-run feasibility of the demand (condition i, Theorem 2) is a necessary and sufficient condition for inducing ergodicity. We show that, in the class of admissible policies (5), hedging-point policy (17) is optimal.

It is worth noting that the proposed policy in (17) is unique in the sense that all other optimal policies with structure (5) are almost everywhere identical to this policy. By contrast, in [3], there is no proof of uniqueness for the optimal hedging-point policy over the specific class it is defined on. This is because the authors establish optimality by using a verification theorem, which constitutes only a sufficient condition, derived from the HJB equation. Moreover, the optimality of the current solution holds for a class of policies slightly broader than that considered in [3] (additional requirements 29.c and 29.d are made in [3]).

Finally, note that, had we attempted to prove the optimality of (17) via a verification theorem on the HJB equation, as in [3], we would have run into serious difficulties. Indeed, part of the challenge is that, for our candidate optimal policy $r(x)$, as in (17), we would have had to display two continuously differentiable functions W_0 and W_1 , and a constant J^* such that

$$[r(x)-d]\frac{\mathrm{d}W_i}{\mathrm{d}x}-q_i[W_i(x)-W_{1-i}(x)] + c(x)-J^*=0 \quad i=0,1, \text{ for all } x$$

where $c(x)$ is the integrand of the long-run expected average criterion. Unfortunately, in our case,

$$c(x)=I_{(l,+\infty)}(c^+x^+ + c^-x^-),$$

which unlike the cost in [3], is not continuous, thus rendering verification impossible.

There are three interesting observations about the optimal hedging-point policy obtained under the current probabilistic criterion (3), in comparison to the one obtained for the long-term expected average cost in [3],

i In contrast to the optimal hedging point obtained in [3], hedging point $z=l/c^+$ in (17) is *always positive*. Thus, unlike [3], a zero-inventory policy for criterion (3) and finite c^+ is never optimal; note however that, as the threshold l approaches zero, the inventory level also approaches zero.

ii The optimal long-run probability (3) is

$$J^*=\frac{k}{k-d}\frac{q_1}{q_0+q_1}\exp\left[-\lambda l\left(\frac{1}{c^-}+\frac{1}{c^+}\right)\right]$$

where

$$\lambda=\frac{q_0}{d}-\frac{q_1}{k-d}>0.$$

Thus as c^- or c^+ go to infinity, probability J^* will not reach its upper bound of 1. In fact, in general

$$J^*\le\frac{k}{k-d}\frac{q_1}{q_0+q_1},$$

where the upper bound is independent of l, c^- and c^+. By contrast, the cost criterion considered in [3] tends to infinity as c^- or c^+ goes to infinity.

iii Hedging point $z=l/c^+$ depends only on the instantaneous holding cost c^+ and threshold l, and unlike that in [3], is *completely independent* of c^-. The instantaneous backlog cost c^- affects only the minimal probability J^* of exceeding instantaneous cost l.

References

[1] Ahmed, S., U. Cakmak and A. Shapiro, 2007, "Coherent risk measures in inventory problems", *European Journal of Operational Research*, 182, 226-238.

[2] Akella, R. and P.R. Kumar, 1986, "Optimal control of production rate in a failure-prone manufacturing system", *IEEE Transactions on Automatic Control*, 31, 116-126.

[3] Bielecki, T.R. and P.R. Kumar, 1988, "Optimality of zero-inventory policies for unreliable manufacturing systems", *Operations Research*, 36, 532-541.

[4] Brémaud, P., R.P. Malhamé and L. Massoulié, 1997, "A manufacturing system with general stationary failure process: Stability and IPA of hedging control policies", *IEEE Transactions on Automatic Control*, 42, 155–170.

[5] Chen, X., M. Sim, D. Simchi-Levi and P. Sun, 2007, "Risk aversion in inventory management", *Operations Research*, 55, 828-842.

[6] Choi, T.M., C.H. Chiu and P. Fu, 2011, "Periodic review multi-period Inventory control under a mean-variance optimization objective", *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, 41, 678-682.

[7] Kimemia, J. and S.B. Gershwin, 1983, "An algorithm for the computer control of a flexible manufacturing system", *IEEE Transactions on Automatic Control*, 15, 353–362.

[8] Malhamé, R.P., 1993, "Ergodicity of hedging control policies in single-part multiple-state manufacturing systems", *IEEE Transactions on Automatic Control*, 38, 340-343.

[9] Sethi, S.P., H. Yan and H. Zhang, 2002, "Optimal and hierarchical controls in dynamic stochastic manufacturing systems: A survey", *Manufacturing and Service Operations Management*, 4, 133-170.