

Smoothing in Servo Processes
Author(s): Harlan D. Mills
Source: *SIAM Review*, Vol. 3, No. 2 (Apr., 1961), pp. 131-139
Published by: Society for Industrial and Applied Mathematics
Stable URL: http://www.jstor.org/stable/2027388
Accessed: 09/09/2011 14:02

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
http://www.jstor.org/page/info/about/policies/terms.jsp

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Society for Industrial and Applied Mathematics is collaborating with JSTOR to digitize, preserve and extend access to *SIAM Review*.

http://www.jstor.org

SIAM Review
Vol. 3, No. 2, April, 1961

SMOOTHING IN SERVO PROCESSES[1]

Harlan D. Mills[2]

1. introduction

We abstract a recurring problem of designing a decision process—a servo mechanism—for a stochastic environment. The motivation arises in inventory theory, guidance and tracking systems, vehicle convoy operations, and other such situations. Our specific interest is in designing servos which keep a system near some "goal" condition in steady state operation, while it is subjected to stochastic influences, and beyond that, which maintain as much dynamic stability as possible. Arrow, Karlin and Scarf [1], Brown [3], Pinkham [7], Vassian [8], and Wiener [9], have been concerned with closely related problems.

For example, an inventory system may satisfy a material conservation relation

$$i_{t+1} = i_t + r_{t-T} - s_t$$

for inventories i_t, requisitions r_t (with delivery cycle T), and sales s_t. It may be required to design a requisition policy such that some average level of inventory is maintained when sales are randomly distributed and a stable pattern of requisitions result. As another illustration, a tracking system may satisfy an error conservation relation

$$e_{t+1} = e_t + a_{t-T} + (d_t - o_t)$$

for errors of location e_t, tracking adjustments a_t (lagged T periods), observed displacements d_t and errors of observation o_t. It may be required to maintain, in steady state operation, some low-level of average error of location while smoothing tracking adjustments as much as possible (to minimize dynamic stresses on a physical system).

In generic terms, we will be interested in a discrete system satisfying

$$e_{t+1} = e_t + c_{t-T} + n_t$$

for errors e_t with servo corrections c_t (lagged T periods), and system noise n_t. The noise is then to be a random process, though, as the examples above indicate, with mean not necessarily zero. We consider arbitrary functions S, of past information, as servos,

$$c_t = S(\cdots, n_{t-2}, e_{t-1}, c_{t-1}, n_{t-1}),$$

and seek servos which maintain low levels of variability in both e_t and c_t. We shall find that these variabilities cannot both be reduced indefinitely, and that a simple class of linear servos is optimal in this respect. Furthermore, this optimal class has alternative representations so that applications are possible where not

[1] Received by the editors January 9, 1961.

[2] Mathematica, Princeton, N. J. This work was supported by the Bureau of Supplies and Accounts, U.S. Navy.

all past information is available. For example, one representation is free of past e_t—a quantity possibly not directly observable in the tracking process above.

In another place [6], these same questions are formulated for the system with a nonlinear, nonconservative, relation

$$e_{t+1} = \max\{M, e_t + c_{t-T} + n_t\},$$

where M represents a "stop"—in inventory theory the restriction to non-negative inventories, in other contexts physical bounds. The results below are shown there to hold asymptotically as M is decreased.

2. DISCRETE SERVO-STOCHASTIC PROCESSES

We define a discrete servo-stochastic process to be a sequence of random variables [2, 4, 5]

$$P = (e_1, c_1, n_1, e_2, c_2, n_2, \cdots),$$

where we interpret

e_t as an "error" in an operation at the beginning of a time period t,

c_t as a "correction" made in the operation at the beginning of period t,

n_t as a "noise" entering the operation during period t,

such that P, in conjunction with a sequence of numbers,

$$I = (\cdots, e_{-1}, c_{-1}, n_{-1}, e_0, c_0, n_0),$$

called the "initial conditions," is characterized, for $t = 0, 1, 2, \cdots$, by the following properties:

$$e_{t+1} = e_t + c_t + n_t, \tag{1}$$

$$c_{t+1} = S(\cdots, n_{t-1}, e_t, c_t, n_t), \tag{2}$$

$$\text{Prob}\{n_t \leqq x\} = N(x), \tag{3}$$

where S is an arbitrary function and N is an arbitrary distribution. Equation (1) represents a conservation of error in the operation (holding with probability 1), (2) specifies a "servo" or "decision policy" S (holding with probability 1), and (3) describes noise—taken to be identically and independently distributed period by period.

It is clear that conditions (1), (2), and (3), with the initial conditions I, uniquely determine the process P, for (1), (2), (3) can be applied recursively to determine the distribution of any partial sequence

$$P_k = (e_1, c_1, n_1, e_2, c_2, n_2, \cdots, e_k, c_k, n_k).$$

Means and variances in P will be written as

$$\bar{e}_t = E(e_t), \qquad \bar{c}_t = E(c_t), \qquad \bar{n}_t = E(n_t),$$

$$\sigma_e^2(t) = E[(e_t - \bar{e}_t)^2], \text{ etc.}$$

P is said to be a stable process if all first and second order moments approach definite limits in time; we define

$$\bar{e} = \lim_{t\to\infty} \bar{e}_t\,, \qquad \bar{c} = \lim_{t\to\infty} \bar{c}_t\,, \qquad \bar{n} = \bar{n}_t\,,$$

$$\sigma_e^{\,2} = \lim_{t\to\infty} \sigma_e^{\,2}(t), \text{ etc.}$$

Assume, hereafter, that $\sigma_n^{\,2} > 0$. We shall be concerned only with stable processes—excluding, for example, servos which allow the error to become unbounded. In stable processes, consider two ratios—"error to noise" and "correction to noise"—defined as

$$k_e = \frac{\sigma_e}{\sigma_n}, \qquad k_c = \frac{\sigma_c}{\sigma_n}. \tag{4}$$

In many circumstances, it is desired to design a servo to keep both of these ratios small—the first gives a measure of "how well the servo is doing," and the second of "how hard the servo is working."

First, a "smoothing capacity" is developed for discrete servo-stochastic processes. This capacity, for any servo, is given by

$$k_e \geqq \frac{1}{2}\left(k_c + \frac{1}{k_c}\right), \tag{5}$$

or, in another form

$$k_e k_c \geqq \tfrac{1}{2}(1 + k_c^{\,2}) > \tfrac{1}{2}, \tag{6}$$

which demonstrates that both these ratios cannot be decreased indefinitely by better servo designs.

Next, a complete optimal class of servos is determined which achieves the equality of (5). These optimal servos turn out to be linear, namely, of the form

$$c_t = \gamma - \alpha(e_t - \epsilon), \tag{7}$$

where $0 < \alpha \leqq 1$, and γ, ϵ are arbitrary. Alternative representations exist which can be used when the errors e_t are not directly observable. That so simple a class of policies is complete and optimal, seems quite fortuitous, considering the fact that they are in competition with all possible servos with bounded mean errors—nonlinear, discontinuous, or what have you.

Finally, we formulate and characterize the effect of information time delays and/or servo-mechanical response times on these ratios and on the smoothing capacity. With a composite time delay of T periods, the smoothing capacity is extended to

$$k_e k_c \geqq \tfrac{1}{2}\sqrt{(1 + k_c^{\,2})^2 + Tk_c^{\,2}}, \tag{8}$$

and the class of policies given by (7) remains the complete optimal class.

3. SMOOTHING THEOREMS

Two theorems follow which develop our main results on smoothing in servo-stochastic processes in the absence of information or servo-mechanical delays. Theorem 1 (Smoothing Capacity) determines the boundary of (5) and thereby the smoothing capacity of servos (with no information or mechanical delay). Theorem 2 (Optimal Servo Class) establishes the completeness and optimality of the class of servos given by (7). Other equivalent forms of this servo class are also presented.

THEOREM 1. (SMOOTHING CAPACITY). *In a stable discrete servo-stochastic process, for any servo whatsoever,*

$$k_e \geqq \frac{1}{2}\left(k_c + \frac{1}{k_c}\right). \tag{5}$$

PROOF. By (1)

$$e_{t+1} = e_t + c_t + n_t$$

and

$$\bar{e}_{t+1} = \bar{e}_t + \bar{c}_t + \bar{n}_t \ ;$$

thus,

$$(e_{t+1} - \bar{e}_{t+1}) = (e_t - \bar{e}_t) + (c_t - \bar{c}_t) + (n_t - \bar{n}_t).$$

Square both sides of this latter equation, take expectations, and then limits to obtain

$$\sigma_e^{\,2} = \sigma_e^{\,2} + \sigma_c^{\,2} + \sigma_n^{\,2} + 2\gamma_{ec}\sigma_e\sigma_c \ ,$$

where γ_{ec} is the correlation between e_t and c_t as $t \to \infty$, $-1 \leqq \gamma_{ec} \leqq 1$, and the other two cross product terms of the right side are zero since n_t is independent of e_t and c_t . Using the definitions of k_e and k_c , this last equation can be rewritten as

$$[k_c^{\,2} + 1 + 2\gamma_{ec}k_ek_c]\sigma_n^{\,2} = 0,$$

which can be rearranged, since $\sigma_n^{\,2} \neq 0$, as

$$-\gamma_{ec}k_e = \frac{1}{2}\left(k_c + \frac{1}{k_c}\right). \tag{9}$$

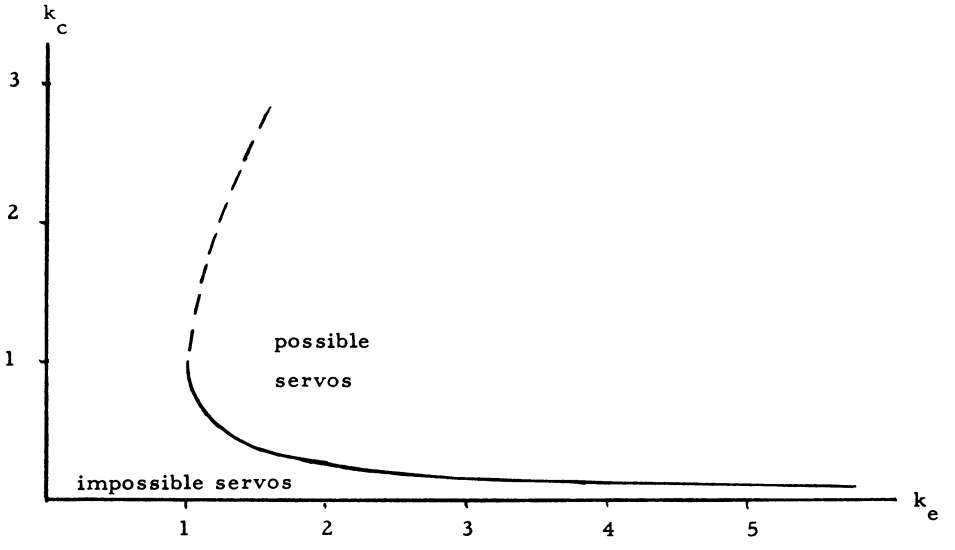
Now $-\gamma_{ec} > 0$ necessarily since $k_e > 0$ and the right side of this equation is positive for any $k_c > 0$. Furthermore, $-\gamma_{ec} \leqq 1$, so

$$k_e \geqq -\gamma_{ec}k_e = \frac{1}{2}\left(k_c + \frac{1}{k_c}\right),$$

as was to be proved.

The boundary provided by Theorem 1 is diagrammed below. We have drawn the boundary of interest as a solid curve, the remainder as a dotted curve. The solid curve represents the barrier to the joint minimization of the ratios k_e and

k_c. Notice that everywhere on this barrier $k_c = \alpha k_e$ for some α such that $0 < \alpha \leqq 1$. Theorem 2 will establish an optimal class of servos which generates this entire barrier.



It is an interesting side point to note that given a certain error to noise ratio k_e, say, not only is there a lower bound to possible correction to noise ratios k_c, but there is an upper bound as well—and these lower and upper bounds are reciprocals. Thus, if servos can only do so well in this respect, they can also only do so badly, and still provide a stable process.

Curiously enough, it will be apparent, from the proof of the next theorem that this "worst possible" servo has the same form as the optimal class, but for a different range of the parameter α—namely, the class given by (7), except that $1 \leqq \alpha < 2$. Thus all servos which provide a given error to noise necessarily have correction to noise ratios which lie between a certain number and its reciprocal.

THEOREM 2. (OPTIMAL SERVO CLASS). *Given α such that $0 < \alpha \leqq 1$, the servo*

$$c_t = \gamma - \alpha(e_t - \epsilon), \tag{7}$$

where γ and ϵ are arbitrary constants, determines a stable discrete servo-stochastic process such that

$$k_e = \frac{1}{2}\left(k_c + \frac{1}{k_c}\right) \quad \textit{and} \quad k_c = \alpha k_e\,. \tag{5'}$$

PROOF. We establish, first, that (7) determines a stable process, and then that it has the required properties. To accomplish this we exhibit a solution to the

pair of difference equations

$$e_{t+1} = e_t + c_t + n_t, \qquad t = 0, 1, 2, \cdots, \tag{1}$$

$$c_{t+1} = \gamma - \alpha(e_{t+1} - \epsilon), \tag{7}$$

with e_0, c_0, n_0 being given numbers. Substitute (7) in (1) to obtain

$$e_{t+1} = (1 - \alpha)e_t + \gamma + \alpha\epsilon + n_t$$

with an easily verified solution

$$e_t = \left[\frac{\gamma}{\alpha} + \epsilon\right][1 - (1 - \alpha)^t] + \sum_{j=0}^{t-1} (1 - \alpha)^j n_{t-1-j} + (1 - \alpha)^t e_0, \qquad t = 1, 2, \cdots, \tag{10}$$

and on resubstituting (10) into (7),

$$c_t = -\alpha \sum_{j=0}^{t-1} (1 - \alpha)^j n_{t-1-j} - \alpha(1 - \alpha)^t e_0 + (\gamma + \alpha\epsilon)(1 - \alpha)^t. \tag{11}$$

Now $|1 - \alpha| < 1$ when $0 < \alpha \leqq 1$ and (10) and (11) certainly exhibit a stable process, since the noise, n_j, is identically and independently distributed.

We notice further, from (7), that

$$\bar{c}_{t+1} = \gamma - \alpha(\bar{e}_{t+1} - \epsilon),$$

so that

$$(c_t - \bar{c}_t) = -\alpha(e_t - \bar{e}_t). \tag{12}$$

Squaring both sides of (12) and taking expectations and limits gives

$$\sigma_c^2 = \alpha^2\sigma_e^2, \quad \text{or} \quad \sigma_c = \alpha\sigma_e,$$

and, hence,

$$k_c = \alpha k_e.$$

On the other hand, multiplying both sides of (12) by $(e_t - \bar{e}_t)$ and taking expectations and limits gives

$$\gamma_{ec}\sigma_e\sigma_c = -\alpha\sigma_e^2 \quad \text{or} \quad \gamma_{ec}\sigma_e = -\alpha\sigma_e,$$

and, hence, since $\sigma_e > 0$,

$$\gamma_{ec} = -1.$$

But referring to Equation (9) in the proof of Theorem 1, we find that if $\gamma_{ec} = -1$, then

$$k_e = \frac{1}{2}\left(k_c + \frac{1}{k_c}\right),$$

as was to be proved. Thus we have shown that (7) determines a stable process and leads to both properties stated in (5′). This completes the proof of the theorem.

Incidentally, the explicit forms of e_t and c_t in (10) and (11) allow us to compute $\bar{e}$, $\bar{c}$, σ_e^2, σ_c^2 as follows:

$$\bar{e} = \epsilon + \frac{1}{\alpha}(\gamma + \bar{n}), \qquad \bar{c} = -\bar{n},$$

$$\sigma_e^2 = \sum_{j=0}^{\infty} (1 - \alpha)^{2j}\sigma_n^2 = \frac{1}{1 - (1 - \alpha)^2}\sigma_n^2,$$

$$\sigma_c^2 = \alpha^2 \sum_{j=0}^{\infty} (1 - \alpha)^{2j}\sigma_n^2 = \frac{\alpha^2}{1 - (1 - \alpha)^2}\sigma_n^2,$$

whence

$$k_e^2 = \frac{1}{1 - (1 - \alpha)^2}, \qquad k_c^2 = \frac{\alpha^2}{1 - (1 - \alpha)^2}. \tag{13}$$

Note, in particular, that k_e and k_c are independent of the arbitrary choice of γ and ϵ in (7). In fact, though there are, formally, three parameters α, γ, ϵ, in (7), only two parameters, say α and $\beta = \gamma/\alpha + \epsilon$ determine the process, and only α affects the error to noise and correction to noise ratios.

The optimal servo class given by

$$c_t = \gamma - \alpha(e_t - \epsilon) \tag{7}$$

has, as we have seen, the alternative representation

$$c_t = -\alpha \sum_{j=0}^{t-1} (1 - \alpha)^j n_{t-1-j} - \alpha(1 - \alpha)^t e_0 + (\gamma + \alpha\epsilon)(1 - \alpha)^t, \tag{11}$$

and (eliminating e_{t+1} and e_t from (1) by means of (7)) still another form is

$$c_{t+1} = (1 - \alpha)c_t - \alpha n_t . \tag{14}$$

Representations (7) and (14) provide alternative means of realizing the optimal class, depending on what information is available. Additional insight into the servo of (7) comes from examining

$$\begin{aligned} E(e_{t+1}|e_t) &= E(e_t + c_t + n_t|e_t) \\ &= E(e_t + \gamma - \alpha(e_t - \epsilon) + n_t|e_t) \\ &= (1 - \alpha)e_t + \gamma + \alpha\epsilon + \bar{n} \\ &= (1 - \alpha)e_t + \alpha\bar{e}. \end{aligned}$$

Thus, the servo always seeks, in expectation, to move a fraction α of the way from e_t to the goal $\bar{e}$—another behavioral characteristic of the servo which is independent of the parameters γ and ϵ. On the other hand, in the "worst possible" servo, obtained by letting $1 \leqq \alpha < 2$ in (7), the correction, in expectation, is reflected beyond the goal $\bar{e}$ (since the coefficient of e_t, $1 - \alpha$ is negative).

4. TIME-DEPENDENT SMOOTHING THEOREMS

The following theorems extend the results of Theorems 1 and 2 to situations with time delays in the servo processes. Theorem 3 (Information Delay) formulates the effect of non zero information time delays and/or servo-mechanical response times on servos. Theorem 4 summarizes and organizes the first three theorems into a comprehensive statement to obtain the time-dependent smoothing capacity of (8) and verification of the completeness and optimality of the class of servos given by (7) under any such condition of information delay.

Up to this point, we have assumed that corrections in the process are made instantaneously. Suppose, instead, that servo-mechanical considerations require T periods for a desired correction to be realized. Then (1) should read

$$(1^T) \qquad e_{t+1} = e_t + c_{t-T} + n_t .$$

On the other hand, suppose every item of information in the process is delayed by T periods. Then (2) should read

$$(2^T) \qquad c_{t+1} = S(\cdots, n_{t-T-1}, e_{t-T}, c_{t-T}, n_{t-T}).$$

It is easy to verify, formally, that the two sets of properties (1^T), (2), (3) and (1), (2^T), (3) determine identical stochastic processes. Thus delays in servo-mechanical responses and delays in information, or any combination of them, are equivalent. For convenience, then, we can simply phrase all such delays as information delays. The effects of such delays are characterized in the following theorem.

THEOREM 3. (INFORMATION DELAY). *Let S^T be identical to servo S except that information is delayed T periods in its use by S^T, i.e.,*

$$S^T(\cdots, n_{t-1}, e_t, c_t, n_t) = S(\cdots, n_{t-T-1}, e_{t-T}, c_{t-T}, n_{t-T});$$

then if S determines a stable discrete servo-stochastic process, S^T also determines a stable process with the properties

$$[k_e(S^T)]^2 = [k_e(S)]^2 + T, \qquad k_c(S^T) = k_e(S).$$

PROOF. Let $P = \{e_t, c_t, n_t\}$ be the stable discrete servo-stochastic process determined by S. Then

$$(1) \qquad e_{t-T+1} = e_{t-T} + c_{t-T} + n_{t-T} .$$

Add, now, the T terms, $n_{t-T+1} + \cdots + n_t$, to both sides of this equation, and regroup, as

$$(e_{t-T+1} + n_{t-T+1} + \cdots + n_t) = (e_{t-T} + n_{t-T} + \cdots + n_{t-1}) + c_{t-T} + n_t .$$

Then $P^T = \{e_t{}^T, c_t{}^T, n_t\}$, where

$$e_t{}^T = e_{t-T} + n_{t-T} + \cdots + n_{t-1}, \qquad c_t{}^T = c_{t-T},$$

satisfies (1), and

$$(2^T) \qquad c_{t+1}{}^T = S^T(\cdots, n_{t-1}, e_t, c_t, n_t);$$

hence P^T is the process determined by S^T. Now, it is clear that P^T is stable if P is. In addition the values of σ_e^2 and σ_n^2 may be computed directly from the form of P^T, and are as given. This completes the proof of the theorem.

Finally, we note that the results of the three preceding theorems can be organized into a single comprehensive statement about discrete servo-stochastic processes as follows:

THEOREM 4. *In a stable discrete servo-stochastic process with information time delay T, for any servo whatsoever,*

$$k_e k_c \geqq \tfrac{1}{2}\sqrt{(1 + k_c^2)^2 + Tk_c^2}. \tag{8}$$

The servo

$$c_t = \gamma - \alpha(e_t - \epsilon), \qquad 0 < \alpha \leqq 1, \tag{7}$$

where γ and ϵ are arbitrary constants, determines a stable process such that the equality holds in (8). *Furthermore, the class of servos given by* (7) *generates the entire minimal boundary of* (8) *as α varies from* 0 *to* 1.

REFERENCES

1. ARROW, K. J., KARLIN, S. AND SCARF, H., *Studies in the Mathematical Theory of Inventory and Production*, Stanford, 1958.
2. BARTLETT, M. S., *An Introduction to Stochastic Processes*, Cambridge, 1956.
3. BROWN, W. M., *Analysis of Discrete Linear Systems*, J. Soc. Indust. Appl. Math., 5 (1957), pp. 206–224.
4. DOOB, J. L., *Stochastic Processes*, Wiley, 1953.
5. FELLER, W., *An Introduction to Probability Theory and Its Applications*, Vol. 1, 2nd ed., Wiley, 1957.
6. MILLS, HARLAN D., *Smoothing in Inventory Processes*, Navy Supply System Research—Study I (M. L. Balinski, Ed.), Mathematica Report, Princeton, 1960.
7. PINKHAM, ROGER S., *An Approach to Linear Inventory-Production Rules*, Operations Res., 6 (1958), pp. 185–189.
8. VASSIAN, HERBERT J., *Application of Discrete Variable Servo Theory To Inventory Control*, Operations Res., 3 (1955), pp. 272–282.
9. WIENER, NORBERT, *Extrapolation, Interpolation and Smoothing of Stationary Time Series*, Wiley, New York, 1949.