

## An Extension of the Best Linear Controller to a Polynomial Controller for Non-Gaussian Disturbances

RAPHAEL SIVAN<sup>1</sup>

The problem of finding the *optimal controller* (or *optimal estimator*), in the *mean square* sense for *linear* systems which are disturbed by *Gaussian additive* noise is, by now, a completely solved problem [1], [2]. In fact, the optimal controller turns out to be a *linear* function of the observations.

In the case that the disturbances are *not* Gaussian, we can still derive the best *linear controller*, which, however, will not necessarily<sup>2</sup> be optimal. For this non-Gaussian case we shall extend the best linear controller to a *polynomial controller*. However, while the construction of the best linear controller requires only the knowledge of the *first* and *second moments* of all random disturbances, the construction of the *n*th order polynomial controller requires the knowledge of (and utilizes!) the first  $2n$  moments.

### A REVIEW OF THE BEST LINEAR CONTROLLER

Let a *dynamical system*  $S$  and a *measuring device*  $M$  be characterized by the following equations

$$S: X_{n+1} = aX_n + U_n + V_n \quad (1)$$

$$M: Y_n = X_n + W_n \quad (2)$$

where  $X_n$  is the *state* at time  $n$ ,  $U_n$  is the *input* and  $V_n$  is the *observable output*;  $\{V_n\}_{n=0}^N$  and  $\{W_n\}_{n=0}^N$  are two sequences of *independent random variables* which are also independent of one another. Essentially,  $\{V_n\}_{n=0}^N$  is the disturbance to the system  $S$ , while  $\{W_n\}_{n=0}^N$  is the disturbance to the measuring device  $M$ .  $X_0$ , the *initial state* is also a *random variable* which is *independent* of  $\{V_n\}_{n=0}^N$  and of  $\{W_n\}_{n=0}^N$ .

For simplicity, we shall assume that (1) and (2) are *scalar* equations; that  $V_i=0, \dots, N$  are *zero mean, equally distributed*, random variables with finite variance; likewise,  $W_i=0, 1, \dots, N$  are *zero mean equally distributed*, random variables with finite variance. We also assume  $X_0$  is zero mean and has finite variance.

The dynamical system  $S$  evolves as follows: at time  $j$  the system is at state  $X_j$ . We, however, cannot measure  $X_j$ , but only  $Y_j$ . Based upon  $Y_j$ , and upon the previous measurements  $Y_0, \dots, Y_{j-1}$ , an input  $U_j(Y_0, \dots, Y_j)$  is applied to the system  $S$ . This input, together with the disturbance  $V_j$ , control the system to the new state  $X_{j+1}$  at time  $j+1$ . The procedure is repeated for  $j=0, 1, \dots, N$ .

It is well known that  $\bar{X}_j$ , the best estimate of  $X_j$ , given  $Y_0, \dots, Y_j$ , in the sense that

$$E\{(X_j - \bar{X}_j)^2 | Y_0, \dots, Y_j\}$$

is minimized, is the conditional expectation, *i.e.*,

$$\bar{X}_j = E\{X_j | Y_0, \dots, Y_j\} \quad j = 0, 1, \dots, N.$$

Moreover, it can be shown that  $U_j^*(Y_0, \dots, Y_j)$ , the optimal input at time  $j$ , given  $Y_0, \dots, Y_j$  in the sense that

$$E\left\{\sum_{i=1}^{N+1} X_i^2(X_0, U_0, \dots, U_{j-1}, V_0, \dots, V_{j-1})\right\} \quad (3)$$

is minimized, is

$$U_j^* = -a\bar{X}_j \quad j = 0, 1, \dots, N. \quad (4)$$

For the special case that  $V_i, W_i, i=0, 1, \dots, N$  and  $X_0$  are all *Gaussian* random variables, explicit formulas can be obtained for  $\bar{X}_j$  and  $U_j^*$ , namely

$$\left. \begin{aligned} \bar{X}_j &= \frac{m_{2;X_j}}{m_{2;X_j} + m_{2;W_0}} Y_j \\ U_j^* &= -a\bar{X}_j \end{aligned} \right\} \quad j = 0, \dots, N \quad (5)$$

where we use the following notation

$$m_{k;X_j} = E\{X_j - E(X_j)\}^k \quad j = 0, 1, \dots, N \quad k = 2, 3, \dots \quad (6)$$

and a similar notation for  $v$  and  $w$ .

Further,  $m_{2;X_j}$  can be computed by using the recurrence relation

$$m_{2;X_{j+1}} = a^2 \frac{m_{2;X_j} + m_{2;W_0}}{m_{2;X_j} + m_{2;W_0}} + m_{2;V_0} \quad (7)$$

Note that (7) holds only if the controls  $U_0, \dots, U_N$ , which are applied to the system, are the optimal ones, *i.e.*, are chosen according to (5).

The main feature of the solutions for  $\bar{X}_j$  and  $U_j^*$  (5) is that only the first two moments of the random variables have to be taken into account.

At first sight it might seem strange that in computing  $\bar{X}_j$  in (5) we use  $m_{2;X_j}$ , the *unconditional* second central moment, and not  $E\{(X_j - \bar{X}_j)^2 | Y_0, \dots, Y_j\}$ , the *conditional* second central moment; namely that we do not make use of the fact that at time  $j$ , the outputs  $Y_0, \dots, Y_j$  are already known. A simple calculation, however, will show that for the Gaussian case the conditional and the unconditional second central moment are the same, *i.e.*,

$$m_{2;X_j} = E\{(X_j - E(X_j | Y_0, \dots, Y_j))^2 | Y_0, \dots, Y_j\} \quad (8)$$

Finally we remark that the solution to the problem of finding the best *linear* estimate and best *linear* control for the case where the random variables are not necessarily Gaussian is the same as given by (5)-(7), except that now

- 1)  $\bar{X}_j$  does not stand for the conditional expectation, but for the *minimum variance linear estimate*.
- 2) Eq. (8) is not satisfied now. In fact, although it might be advantageous to use the conditional second central moments, we must use  $m_{2;X_j}$  since the conditional moments will lead to nonlinear controllers.

However, we might not be satisfied with the best *linear* controller for this *non-Gaussian* case. Assuming that more statistical information than merely the first and second moments is available, how can we find a better approximation (necessarily a nonlinear one) to the optimal controller than the best linear controller? The answer to this question is attempted in the next section.

### THE POLYNOMIAL CONTROLLER

Let us assume that not only the *two* first moments, but all *four* first moments of all random variables are known. The question is how to make use of this additional information in order to improve the best linear controller.

A natural extension of the best linear controller (for the system (1), (2) and the cost (3)) would be a controller of the form

$$\left. \begin{aligned} U_j^* &= -a\bar{X}_j \\ \bar{X}_j &= a_j(Y_j - m_j)^2 + b_j(Y_j - m_j) + c_j \end{aligned} \right\} \quad j = 0, \dots, N \quad (9)$$

where

$$m_j = E\{X_j\} \quad (10)$$

and the constants  $a_j, b_j$ , and  $c_j$  are chosen so as to minimize

$$E\{X_j - \bar{X}_j\}^2.$$

The procedure of computing  $a_0, b_0$  and  $c_0$  is straightforward. By equating the partial derivatives of  $E\{X_0 - \bar{X}_0\}^2$ , with respect to  $a_0, b_0$  and  $c_0$  to zero, and solving the linear equations obtained we get the following results:

$$a_0 = \frac{1}{d_0} [m_{3;X_0}(m_{2;X_0} + m_{2;W_0}) - m_{2;X_0}(m_{3;X_0} + m_{3;W_0})] \quad (11)$$

$$b_0 = \frac{1}{d_0} [m_{2;X_0}(m_{4;X_0} + 6m_{2;X_0}m_{2;W_0} + m_{4;W_0}) - m_{2;X_0}(m_{2;X_0} + m_{2;W_0})^2 - m_{3;X_0}(m_{3;X_0} + m_{3;W_0})] \quad (12)$$

$$c_0 = \frac{1}{d_0} [m_{2;X_0}(m_{3;X_0} + m_{3;W_0})(m_{2;X_0} + m_{2;W_0}) - m_{3;X_0}(m_{2;X_0} + m_{2;W_0})^2] + m_0 \quad (13)$$

Manuscript received, January 21, 1964.  
The author is with the Department of Electrical Engineering, California Institute of Technology, Pasadena, Calif.  
<sup>1</sup> Formerly Raphael Sussman.  
<sup>2</sup> In fact, it will never be optimal. This statement is proved by Sivan [3].

where

$$d_0 = (m_{2;X_0} + m_{2;W_0})(m_{4;X_0} + 6m_{2;X} m_{2;W_0} + m_{4;W_0}) - (m_{3;X_0} + m_{3;W_0})^2 - (m_{2;X_0} + m_{2;W_0})^3. \tag{14}$$

Note that  $a_0$ ,  $b_0$  and  $c_0$  depend only upon the first four moments of  $X_0$  and  $W_0$ , so that the knowledge of all higher moments is not necessary.

The second control will be found in a similar fashion as  $U_0^*$ , namely, by formulas similar to (9)–(14), except that  $X_0$  and  $W_0$  are replaced by  $X_1$  and  $W_1$ . Namely, we denote

$$U_1^* = -a\tilde{X}_1 \tag{15}$$

where

$$\tilde{X}_1 = a_1(Y_1 - m_1)^2 + b_1(Y_1 - m_1) + c_1. \tag{16}$$

However, in order to solve for  $a_1$ ,  $b_1$  and  $c_1$ , the four first moments of  $X_1$  will have to be computed. We start now to compute these moments. By (1) and (9)

$$X_1 = aX_0 - a\tilde{X}_0 + V_0 \tag{17}$$

so that

$$m_1 = E\{X_1\} = aE\{X_0\} - aE\{\tilde{X}_0\} + E\{V_0\} = am_0 - aa_0(m_{2;X_0} + m_{2;W_0}) - ac_0 = 0. \tag{18}$$

Next

$$m_{2;X_1} = a^2E\{X_0 - X_0\}^2 + m_{2;V_0} \tag{19}$$

$$m_{3;X_0} = a^3E\{X_0 - X_0\}^3 + m_{3;V_0} \tag{20}$$

$$m_{4;X_1} = a^4E\{X_0 - X_0\}^4 + 6a^2E\{X_0 - X_0\}^2 m_{2;V_0} + m_{4;V_0}. \tag{21}$$

By inserting these moments into (11)–(14), instead of the moments of  $X_0$ , we would be able to solve for the constants  $a_1$ ,  $b_1$  and  $c_1$ ; however, here we encounter a difficulty which distinguishes the present problem from the linear approximation problem dealt with above. The difficulty is that in order to compute  $m_{k;X_1}$  we shall have to know the  $2k$  first moments of  $X_0$ . Further, in order to compute  $m_{k;X_2}$  the  $4k$  first moments of  $X_0$  will have to be known; and so on. Thus a straightforward application of the above presented method would not be possible unless all the moments of  $X_0$  were known. Moreover, even if all the moments were known, we would not be satisfied with that method, since at every step it would require that we perform a different operation upon the observations, whereas we are looking for an algorithm which can be expressed as a relatively simple, time-invariant recurrence relation, where all quantities at the  $k+1$ th step depend only upon the quantities at the  $k$ th step.

We propose the following method to overcome this difficulty: we shall replace  $X_0$  by a random variable  $Z_0$ , where  $Z_0$  is chosen so that

- 1)  $E\{Z_0\} = m_0$
- 2)  $m_{i;Z_0} = m_{i;X_0}$   $i = 2, 3, 4$
- 3) the probability density function of  $Z_0$  belongs to some four parameter family of density functions.<sup>3</sup>

Having chosen such a family of density functions, all the moments of  $Z_0$  can be computed, and it is with these moments that we shall approximate  $m_1$ ,  $m_{2;X_1}$ ,  $m_{3;X_1}$  and  $m_{4;X_1}$ .<sup>4</sup> This completes the computation of  $U_1^*$ . In order to find  $U_2^*$ , we shall have, again, to replace  $X_1$  by  $Z_1$ , where  $Z_1$  is a random variable such that its four first moments are identical to those of  $X_1$  (which were computed approximately, using  $Z_0$ ). We also require that  $Z_1$  be a random variable whose density belongs to the same four parameter family as  $Z_0$ . This assumption allows us to get approximate values for the higher moments of  $X_1$ , and thus compute  $U_2^*$  using (9)–(14).

This process of approximating the random variable  $X_j$ , of which only the first four moments are known, by a random variable  $Z_j$ ,

which has a certain fixed density function except for four adjustable parameters, will be repeated at every step. These four free parameters will be adjusted so that the first four moments of  $X_j$  and of  $Z_j$  are the same. After adjusting these four parameters, the density function is completely defined, and thus the next four moments of  $Z_j$  can be computed.

Finally, we arrive at equations of the following form:

$$m_{Z_{k+1}} = K(m_{Z_k}, m_{2;Z_k}, m_{3;Z_k}, m_{4;Z_k}) Y_k^2 + L(m_{Z_k}, m_{2;Z_k}, m_{3;Z_k}, m_{4;Z_k}) Y_k + M(m_{Z_k}, m_{2;Z_k}, m_{3;Z_k}, m_{4;Z_k}) \tag{22}$$

$$m_{i;Z_{k+1}} = N_i(m_{Z_k}, m_{2;Z_k}, m_{3;Z_k}, m_{4;Z_k}) \quad i = 2, 3, 4 \tag{23}$$

where

$$m_{Z_k} = E\{Z_k\}$$

$$m_{i;Z_k} = E\{Z_k - E(Z_k)\}^i \quad i = 2, 3, 4.$$

The functions  $K$ ,  $L$ ,  $M$  and  $N_i$  will depend upon the specific four parameter density function chosen.

Using the approximate moments  $m_{i;Z_j}$  (instead of the exact ones  $m_{i;X_j}$ ) in (9)–(14) will allow us to get an approximation to the optimal control.

### CONCLUSION

The polynomial controller just presented is an extension of the best linear controller, in the sense that in the case that all random variables are Gaussian, the controller reduces to the best linear controller (provided the four parameter family of density function reduces to the Gaussian distribution).

In case that the random variables are non-Gaussian and some of their higher order moments are known, this polynomial controller will make use of these moments to improve upon the best linear controller.

### APPENDIX

In the Appendix we present an example of a four parameter family of density functions which reduce in a special case to the Gaussian density. This family is called the Pearson system [4].

Let  $p(x)$  be the solution of the differential equation

$$\frac{dp}{dx} = \frac{p(m-x)}{a+bx+cx^2}. \tag{24}$$

The constant of integration can be chosen so that  $p(x)$ , which is a four parameter positive function, will be a probability density. In the special case that  $b=c=0$ ,  $p$  will be the Gaussian density.

Let  $X$  be a random variable whose four first moments are  $m_X$ ,  $m_{2;X}$ ,  $m_{3;X}$ ,  $m_{4;X}$ . If  $X$  is a random variable whose density belongs to the Pearson family, then the parameters are given by

$$c = \frac{\alpha_3}{2(1+2\delta)}, \quad b = \frac{\delta}{2(1+2\delta)}, \quad m = -b, \quad a = 1 - 3c \tag{25}$$

where

$$\alpha_3 = \frac{m_{3;X}}{(m_{2;X})^{3/2}}, \quad \delta = \frac{2\alpha_4 - 6 - 3\alpha_3^2}{\alpha_4 + 3}, \quad \alpha_4 = \frac{m_{4;X}}{m_{2;X}^2}.$$

The higher moments of  $X$  can be computed by the following recurrence relation:

$$\alpha_{n+1} = \frac{m_n \alpha_n - (1 - 3c) \alpha_{n-1}}{c(n+2) - 1}$$

where

$$\alpha_n = \frac{m_{n;X}}{(m_{2;X})^{n/2}}.$$

### REFERENCES

- [1] R. F. Kalman, "A new approach to linear filtering and prediction problems," *ASME Trans., J. of Basic Engr.*, pp. 35-45; March, 1960.
- [2] K. J. Astrom, R. W. Koepecke, and F. Tung, "On the Control of Linear Discrete Dynamic Systems with Quadratic Loss," *IBM Res. Rept.*, no. RJ-222; September 1962.
- [3] R. Sivan, "The Necessary and Sufficient Conditions for the Optimal Controller to Be Linear," *Proc. 1964 Joint Automatic Control Conf.*, Stanford, Calif., pp. 297-304.
- [4] J. F. Kenney and E. S. Keeping, "Mathematics of Statistics," Pt. 2, D. Van Nostrand Co., Inc., New York, N. Y., 2nd. Ed., pp. 101-107; 1951.

<sup>3</sup> An example of such a family is presented in the Appendix. This particular family reduces to the Gaussian family in the case that  $m_{2;X_j} = 0$  and  $m_{4;X_j} = 3m_{2;X_j}^2$ . Thus it is expected that for nearly Gaussian random variables it will serve as a reasonable approximation.

<sup>4</sup> Actually, the values found for  $m_1$  and for  $m_{2;X_1}$  turn out to be the exact ones.