

## EFFECT OF CORRELATED ESTIMATION ERRORS IN ORDINAL OPTIMIZATION

Mei Deng

Division of Applied Sciences  
Harvard University

Y.C. Ho

Division of Applied Sciences  
Harvard University

J.Q. Hu

Department of Manufacturing Eng.  
Boston University

### ABSTRACT

Ordinal Optimization (Ho, Sreenivas, and Vakili 1992) used in discrete event simulation promises to be an efficient tool for narrowing the search for optimum. The theory so far, however, is based on independent estimation error for parametrically different experiments. This paper experimentally verifies the intuitive belief that correlations in estimation error can only help but not hinder the search process. Some analytical results are also presented to support the results. The results support the notion of synergy between ordinal optimization and parallel simulation on a SIMD machine.

### 1. INTRODUCTION AND RATIONALE

In a separate paper (Ho, Sreenivas, and Vakili 1992), we documented the advantages of doing ordinal rather than cardinal optimization. In other words, if we are primarily interested in comparing the relative order of the performance of the various designs vs. the actual magnitude of the performance of the designs, then considerable computational savings can be expected. Principally, we demonstrated both theoretically and experimentally that performance order are largely immune to corrupting noise. This means that we can evaluate system performance very approximately and still have confidence that ordinal comparisons based on such approximate estimate of performance are valid. For example, suppose we randomly pick out 12 designs out of 200 in the space of design parameters. We can lay even bets that at least one of the 12 designs will remain in the top 6% performance population even if we randomly evaluate another 188 designs. (This correspond to the case of estimating and ordering 200 designs where the estimation errors are infinite. For a general solution to this problem see Appendix of Ho, Sreenivas, and Vakili (1992). The seemingly counter-intuitive result is similar in nature to the probability result of birthday matching in a room of 25 person).

One of the assumptions of both the analytical and experimental results of ordinal optimization reported in

Ho, Sreenivas, and Vakili (1992) is that the performance estimation errors are i.i.d. In applications such as performance estimation for DEDS via simulation, this assumption of independent estimation error from one design to next may not hold due to the use of common random variables, replications with identical initial conditions, and parallel simulation (Vakili, Mollamustafaoglu, and Ho 1992), etc. Question then naturally arises as to the validity of the noise immunity property of ordinal optimization in such cases. Intuitively, we expect that correlation among the estimation errors seldom can hurt and actually helps most of the time. In the extreme case of positive correlation, i.e. all estimation errors are the same. It is clear, the observed (estimated) order of performance will coincide with the actual order. Similarly, in the case of extreme negative correlation between adjacent performances, by which we mean that only half of the estimation errors are positively correlated, then the effect at worst is to remove half of the performance samples from consideration and the remaining half will be positively correlated which again helps. The purpose of this short paper is to confirm theoretically and experimentally the above conclusion and the intuitive justification above.

### 2. OUR MODEL

We consider performance measure  $J(\theta)$ , say of a discrete event dynamic system, where  $\theta = \{\theta_1, \dots, \theta_N\} \in \Theta$  is the design parameter(s) of interest. We are interested in the following optimization problem:

$$\text{Min}_{\theta \in \Theta} J(\theta)$$

For notational simplicity, we shall use  $J_n$  to denote  $J(\theta_n)$  ( $n = 1, \dots, N$ ). Let us also assume without loss of generality that the  $N$  designs,  $\{\theta_1, \dots, \theta_N\}$ , are ordered (renumbered if necessary) such that  $J_1$  is the best design,  $J_2$  the next best,  $\dots$ ,  $J_N$  the worst. Of course, this is not known to the designer. His job is to determine the best design by observing the noisy versions of  $J_i$ . Let  $J_{[n]}$  be the  $n$ th observed performance of  $J$ . Thus,  $J_{[n]}$  is an estimate of  $J_n$  and  $w_n$  the estimation error, i.e.,

$J_{[n]}=J_n+w_n$ . Our previous work assumes that the  $w_n$ 's are u.i.i.d. A reasonable model of correlated  $w_n$  can be constructed from well known linear system theory. In particular, we consider the following model

$$\begin{pmatrix} w_1 \\ \vdots \\ w_N \end{pmatrix} = A \begin{pmatrix} v_1 \\ \vdots \\ v_N \end{pmatrix} \quad (1)$$

where  $A$  is an  $N \times N$  matrix and  $v_1, \dots, v_N$  are i.i.d. zero mean normal random variables. This model is general enough for most problems of practical interest. However, it does assume zero mean for the  $w_n$ 's (We already pointed out that a constant mean does not matter in ordinal optimization. But the model of (1) does not account for perverse non-constant mean such as the largest positive mean at the best design and monotonically decreasing to the reverse at the worst design). The normal distribution assumption can be easily justified based on the central limit theorem for most cases. Also this is the only reasonable model which satisfies the stationary condition that the marginal distribution of each  $w$  is the same. In Appendix B, we numerically investigate the covariance matrix  $A$  for some simple queueing systems. As we discover, matrix  $A$  often has some nice and simple properties.

As discussed in Ho, Sreenivas, and Vakili (1992), the types of ordered performance curves, (i.e., a plot of the performance of various designs ordered according to magnitude. Thus, the curve must be monotonically increasing if we number the best design as #1, the next best as #2, . . . etc.) particularly for behaviors near the global optimum are limited to three: steep, flat and linear. To this end we suggest a generic set of such ordered performance curves parameterized by  $k$  according to Equation (2)

$$J_n = \begin{cases} \frac{kC}{(1-k)N} & n < (1-k)N \\ \frac{(1-k)C}{kN}n + \frac{(2k-1)C}{k} & n \geq (1-k)N \end{cases} \quad (2)$$

which is illustrated in Fig.1

### 3. EXPERIMENTAL STUDY

Given the estimation error model and the generic set of performance curve in section 2, we are in a position to do a comprehensive set of experiments and to answer the question

**What is the effect of correlation in performance estimation errors on the noise immunity property of ordinal optimization? (Q1)**

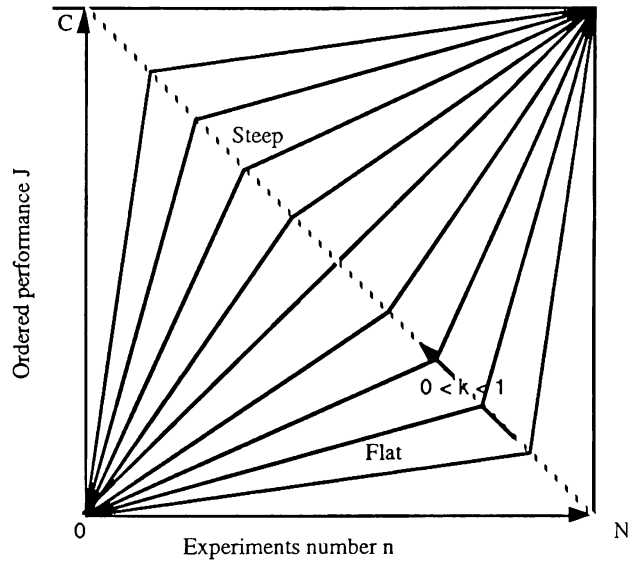


Fig.1 A Generic Set of Ordered Performance Curves

We first consider a simple case in which we assume  $w_1 = v_1$ , and  $w_{n+1} = aw_n + bv_{n+1}$  ( $n = 1, \dots, N-1$ ), where  $v_1, \dots, v_N$  are i.i.d. normal random variables  $N(0,10000/12)$  (Here we assume the noise is relatively large compare to the largest performance value 200. In Ho, Sreenivas, and Vakili (1992), they use  $U(0, 100)$  as their noise model which has mean 50 and variance  $10000/12$ . So we choose the same variance to get the experiment results.) and  $a$  and  $b$  are two constants such that  $a^2+b^2=1$  (this condition insure that the variance of any  $w$  is equal to the variance of  $v$  which is  $10000/12$ ). We know that in this model the noise is exponentially distributed. Matrix  $A$  in (1) is then given by

$$A = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ a & b & 0 & \dots & 0 \\ a^2 & ab & b & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ a^{N-1} & a^{N-2}b & a^{N-3}b & \dots & b \end{pmatrix}$$

We can easily verify that the correlation of estimation errors is increasing with respect to 'a' and geometrically decreasing from one evaluation to the next. In the experiments, we vary

- $k = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9.$
- $a = 0, \pm 0.1, \pm 0.2, \pm 0.3, \pm 0.4, \pm 0.5, \pm 0.6, \pm 0.7, \pm 0.8, \pm 0.9, \pm 0.91, \pm 0.92, \pm 0.93, \pm 0.94, \pm 0.95, \pm 0.96, \pm 0.97, \pm 0.98, \pm 0.99, \pm 1.$

and set  $N = 200, C = 200$ . The number of replications of the experiment for each combination of "a" and "k" is 1100.

We answer a specific version of (Q1)

**What is the average number of observed top-ten designs remaining as true top-ten, i.e. the alignment between observed and actual performance order? (Q1S)**

We summarize our experimental results in Fig.2.

It can be observed from the figure that

- When the correlated factor “ $|\alpha|$ ” increases from 0 to 1, the average number of good designs,  $m$ , does not decrease. When  $0 \leq \alpha \leq 0.9$ ,  $m$  is almost unchanged. When  $\alpha > 0.9$ ,  $m$  increases quickly as to be expected. So we can say the correlation definitely does not make our results worse.

- In the extreme case of positive correlation, i.e.  $w_i = w_j$  for  $i, j$ , the observed order of performance will coincide with the actual order. In the extreme case of negative correlation between adjacent performances, i.e.  $w_{i+1} = -w_i$ , the number  $m$  is always more than 5 but less than 10. This confirms the intuition that the effect at worst is to remove half of the performance samples from consideration. This effect is also seen in the figure.

- When “ $k$ ” increases, i.e. the curve becomes steeper, the number of good designs also increases, i.e. the alignment approaches one. But it is a nonlinear increase. So the steeper the performance curve is, the better the result is.

Generally, for an arbitrary system the noise model may not be exponentially distributed or stationary, (there are some examples in appendix B.) i.e.  $a$  and  $b$  may not be constants. So we did some experiments for a more

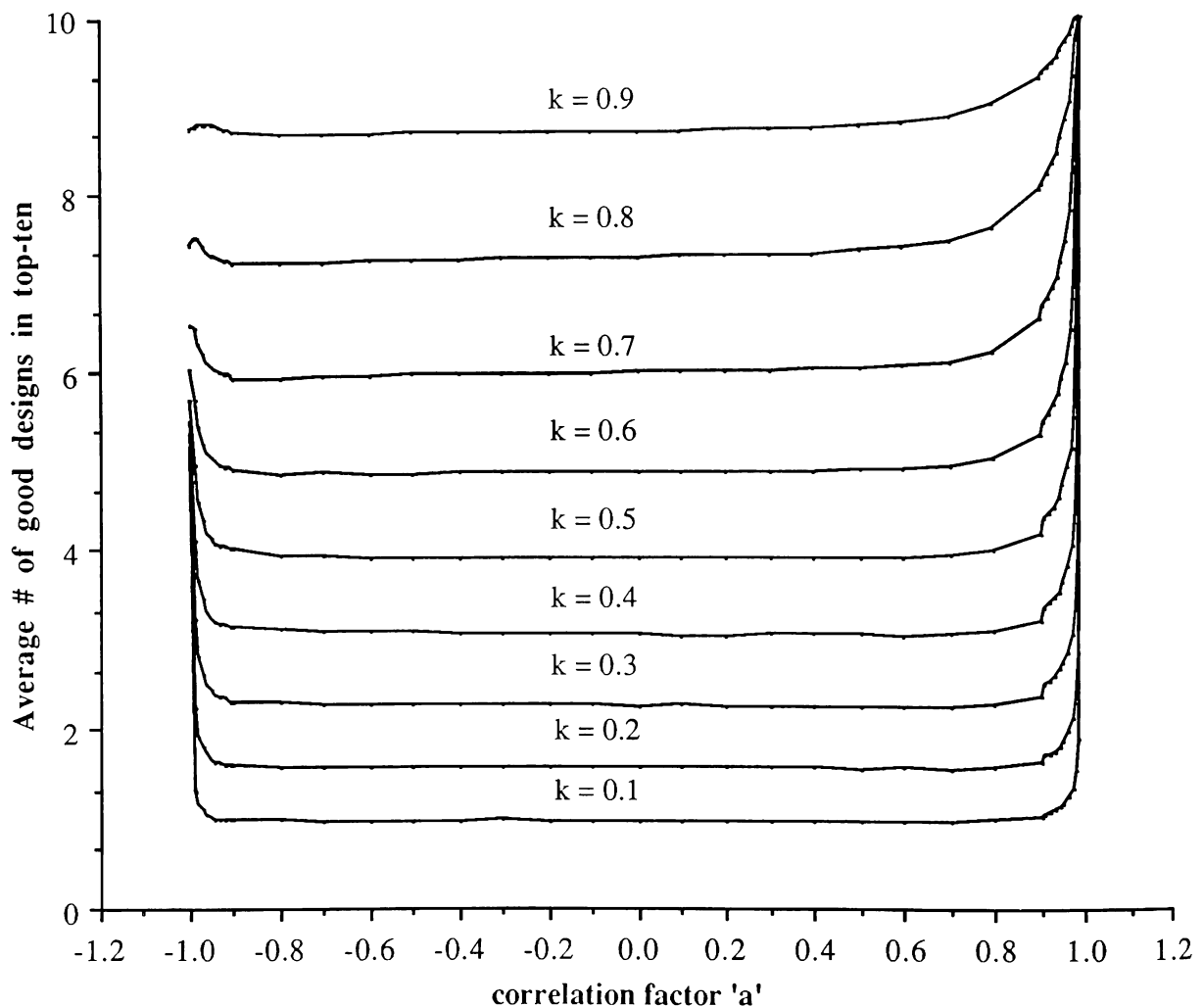


Figure 2. Experimental Results for Normal Noise

complicated system in which  $w_1 = v_1$  and  $w_{n+1} = ca_{n+1}w_n + b_{n+1}v_{n+1}$ . Again similar the previous case  $v_1, \dots, v_N$  are i.i.d. normal random variables  $N(0,10000/12)$ .  $c$  is a scalar constant in  $[0, 1]$ .  $a_n$ 's and  $b_n$ 's are arbitrary non-negative constants in  $[0, 1]$ .  $a_n$ 's represent the relative ratios of the correlations while  $c$  represents the scale of the correlations.  $c, a_n$ 's and  $b_n$ 's satisfy  $(ca_n)^2 + (b_n)^2 = 1$ . Matrix  $A$  in (1) is given by

$$A = \begin{pmatrix} 1 & 0 & \dots & c \\ a_1 c & b_1 & \dots & 0 \\ a_2 a_1 c^2 & a_2 c b_1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ \prod_{i=1}^{N-1} a_i c^{N-1} & \prod_{i=2}^{N-1} a_i c^{N-2} b_i & \dots & b_{N-1} \end{pmatrix}$$

Clearly, the previous case is a special case of this one in which  $a_n = 1$  for all  $n$  and  $c = a$ . Here we vary  $c$  and it can be easily to verify that the correlation is increasing with respect to  $c$ . The experimental results we obtained are similar to those of the previous case.

Before closing this section, we mention that we also tested the case in which  $v_1, \dots, v_N$  are i.i.d. uniform random variables (we want to emphasize that in this case  $w_n$ 's do not have the same type of distribution). Our results for this case have been included in Appendix A, in which we can see that the very similar, though slightly different, conclusions to those we observed for the normal distribution case can be drawn.

#### 4. THEORETICAL ANALYSIS

In the previous section, we show experimentally how correlated noise can affect ordinal optimization. In this section, we will provide some theoretical analysis to back up some of our claims and observations we made based on experimental results.

We first consider the case  $N=2$ . If  $(w_1, w_2)$  is normally distributed with mean and covariance matrix

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$$

respectively, then it is not difficult for us to calculate

$$\text{Prob}(J_{[1]} < J_{[2]}) = \frac{1}{\sqrt{2\pi(\sigma_{11} + \sigma_{22} - 2\sigma_{12})}} \int_{-\infty}^d \text{Exp}\left(-\frac{x^2}{2(\sigma_{11} + \sigma_{22} - 2\sigma_{12})}\right) dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{d/\sqrt{\sigma_{11} + \sigma_{22} - 2\sigma_{12}}} \text{Exp}\left(-\frac{x^2}{2}\right) dx$$

where  $d = J_2 - J_1 \geq 0$ . It is clear  $\text{Prob}(J_{[1]} < J_{[2]})$  is an increasing function with respect to  $\sigma_{12}$ . This means that the larger  $\sigma_{12}$  is, the more likely we will pick up the right design — number one.

We now consider the case  $N=3$ . In this case, we examine the effect of  $\sigma_{23}$  on  $\text{Prob}(J_{[1]} < J_{[2]}, J_{[1]} < J_{[3]})$ , which is the probability that the only design we pick up based on our estimation is the top one. Again, we assume that  $(w_1, w_2, w_3)$  is normally distributed with mean and covariance matrix

$$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{pmatrix}$$

respectively. Note

$$\text{Prob}(J_{[1]} < J_{[2]}, J_{[1]} < J_{[3]}) = \text{Prob}(w_1 - w_2 < c_1, w_1 - w_3 < c_2),$$

where  $c_1 = J_2 - J_1 \geq 0$  and  $c_2 = J_3 - J_1 \geq 0$ . Also note  $(w_1 - w_2, w_1 - w_3)$  is normal distributed with mean and covariance matrix

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \sigma_{11} + \sigma_{22} - 2\sigma_{12} & \sigma_{11} + \sigma_{23} - \sigma_{12} - \sigma_{13} \\ \sigma_{11} + \sigma_{23} - \sigma_{12} - \sigma_{13} & \sigma_{11} + \sigma_{33} - 2\sigma_{13} \end{pmatrix}.$$

Therefore, according to the following lemma we know that  $\text{Prob}(J_{[1]} < J_{[2]}, J_{[1]} < J_{[3]})$  is an increasing function with respect to  $\sigma_{23}$ .

**Lemma 1.** If  $(\xi_1, \xi_2)$  is normally distributed with mean and covariance matrix

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \rho_{11} & \rho_{12} \\ \rho_{12} & \rho_{22} \end{pmatrix}$$

respectively, then  $\text{Prob}(\xi_1 < c_1, \xi_2 < c_2)$  is an increasing function with respect to  $\rho_{12}$  for  $c_1 \geq 0$  and  $c_2 \geq 0$ .

*Proof.* First we have

$$\begin{aligned} \text{Prob}(\xi_1 < c_1, \xi_2 < c_2) &= \frac{1}{2\pi\sqrt{\rho_{11}\rho_{22}-\rho_{12}^2}} * \\ &\int_{x_1 \leq c_1} \int_{x_2 \leq c_2} \text{Exp}\left(-\frac{1}{2}(x_1 \ x_2) \begin{pmatrix} \rho_{11} & \rho_{12} \\ \rho_{12} & \rho_{22} \end{pmatrix}^{-1} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right) dx_1 dx_2 \end{aligned}$$

$$= \frac{1}{2\pi} \int_{\sqrt{\rho_{11}y_1} \leq c_1} \text{Exp}\left(-\frac{1}{2}(y_1^2+y_2^2)\right) dy_1 dy_2, \quad (3)$$

$$\rho_{12}y_1 + \sqrt{\rho_{11}\rho_{22} - \rho_{12}^2}y_2 \leq \sqrt{\rho_{11}c_2}$$

where in the last equality we have made the following variable change

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \sqrt{\rho_{11}} & 0 \\ \rho_{12}/\sqrt{\rho_{11}} & \sqrt{(\rho_{11}\rho_{22} - \rho_{12}^2)/\rho_{11}} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

For notational simplicity, we denote

$$A(\rho_{12}) \equiv \{(y_1, y_2): \sqrt{\rho_{11}}y_1 \leq c_1, \rho_{12}y_1 + \sqrt{\rho_{11}\rho_{22} - \rho_{12}^2}y_2 \leq \sqrt{\rho_{11}c_2}\}$$

and

$$B(\rho_{12}) \equiv \{(y_1, y_2): \rho_{12}y_1 + \sqrt{\rho_{11}\rho_{22} - \rho_{12}^2}y_2 \leq \sqrt{\rho_{11}c_2}\}.$$

In what follows we shall show that the integral in (3) increases with respect to  $\rho_{12}$ , which is equivalent to show that the area  $A(\rho_{12})$  becomes *relatively* larger in terms of the integrand in (3) when  $\rho_{12}$  increases. Let us first drop the condition  $\sqrt{\rho_{11}}y_1 \leq c_1$  in (3) consider

$$\frac{1}{2\pi} \int_{\rho_{12}y_1 + \sqrt{\rho_{11}\rho_{22} - \rho_{12}^2}y_2 \leq \sqrt{\rho_{11}c_2}} \text{Exp}\left(-\frac{1}{2}(y_1^2+y_2^2)\right) dy_1 dy_2, \quad (4)$$

Note that the distance from the origin (0,0) to the line  $\rho_{12}y_1 + \sqrt{\rho_{11}\rho_{22} + \rho_{12}^2}y_2 = \sqrt{\rho_{11}c_2}$  is equal to  $c_2/\sqrt{\rho_{22}}$ , which is independent of  $\rho_{12}$ , hence the line  $\rho_{12}y_1 + \sqrt{\rho_{11}\rho_{22} + \rho_{12}^2}y_2 = \sqrt{\rho_{11}c_2}$  rotates around the origin (0,0) clock-wise as  $\rho_{12}$  increases, which implies that the integral in (4) is independent of  $\rho_{12}$ . Suppose  $\rho_{12}' \leq \rho_{12}$ . Let  $B_1 \equiv B(\rho_{12}) - B(\rho_{12}')$ ,  $B_2 \equiv B(\rho_{12}') - B(\rho_{12})$ , i.e.,  $B(\rho_{12}) = B_1 + B(\rho_{12}) \cap B(\rho_{12}')$  and  $B(\rho_{12}') = B_2 + B(\rho_{12}) \cap B(\rho_{12}')$  (see Figure 3).

Since

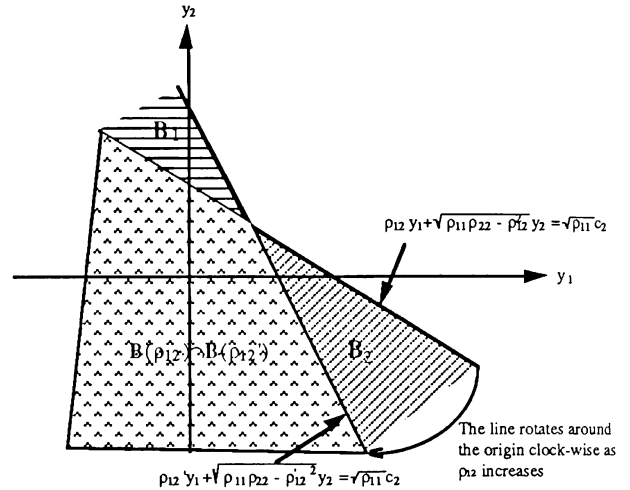
$$\frac{1}{2\pi} \int_{\rho_{12}y_1 + \sqrt{\rho_{11}\rho_{22} - \rho_{12}^2}y_2 \leq \sqrt{\rho_{11}c_2}} \text{Exp}\left(-\frac{1}{2}(y_1^2+y_2^2)\right) dy_1 dy_2$$

$$= \frac{1}{2\pi} \int_{\rho_{12}'y_1 + \sqrt{\rho_{11}\rho_{22} - \rho_{12}'^2}y_2 \leq \sqrt{\rho_{11}c_2}} \text{Exp}\left(-\frac{1}{2}(y_1^2+y_2^2)\right) dy_1 dy_2,$$

we have

$$\frac{1}{2\pi} \int_{B_1} \text{Exp}\left(-\frac{1}{2}(y_1^2+y_2^2)\right) dy_1 dy_2$$

$$= \frac{1}{2\pi} \int_{B_2} \text{Exp}\left(-\frac{1}{2}(y_1^2+y_2^2)\right) dy_1 dy_2, \quad (5)$$



$$B(\rho_{12}) = B_1 + B(\rho_{12}) \cap B(\rho_{12}')$$

$$\text{and } B(\rho_{12}') = B_2 + B(\rho_{12}) \cap B(\rho_{12}')$$

Figure 3: Definition of  $B_1$  and  $B_2$

We also note that since the line  $\rho_{12}y_1 + \sqrt{\rho_{11}\rho_{22} + \rho_{12}^2}y_2 = \sqrt{\rho_{11}c_2}$  rotates around the origin (0,0) clock-wise as  $\rho_{12}$  increases,  $B_1$  must be located at the left-hand side of  $B_2$ . Therefore if we define  $A_1 \equiv B_1 \cap \{\sqrt{\rho_{11}}y_1 \leq c_1\}$ , and  $A_2 \equiv B_2 \cap \{\sqrt{\rho_{11}}y_1 \leq c_1\}$ , then we have either (1)  $A_1 = B_1$  or (2)  $A_2 = 0$ , depending on whether the intersection of two lines  $\rho_{12}y_1 + \sqrt{\rho_{11}\rho_{22} + \rho_{12}^2}y_2 = \sqrt{\rho_{11}c_2}$  and  $\rho_{12}'y_1 + \sqrt{\rho_{11}\rho_{22} + \rho_{12}'^2}y_2 = \sqrt{\rho_{11}c_2}$  is on the left-hand side or right-hand side of the line  $\sqrt{\rho_{11}}y_1 = c_1$  (see Figures 4a and 4b). Based on (5), we have in either case

$$\frac{1}{2\pi} \int_{A_1} \text{Exp}\left(-\frac{1}{2}(y_1^2+y_2^2)\right) dy_1 dy_2$$

$$\geq \frac{1}{2\pi} \int_{A_2} \text{Exp}\left(-\frac{1}{2}(y_1^2+y_2^2)\right) dy_1 dy_2,$$

Since  $A(\rho_{12}) = A_1 + A(\rho_{12}) \cap A(\rho_{12}')$  and  $A(\rho_{12}') = A_2 + A(\rho_{12}) \cap A(\rho_{12}')$ , we have

$$\frac{1}{2\pi} \int_{A_1} \text{Exp}\left(-\frac{1}{2}(y_1^2+y_2^2)\right) dy_1 dy_2$$

$$\geq \frac{1}{2\pi} \int_{A_2} \text{Exp}\left(-\frac{1}{2}(y_1^2+y_2^2)\right) dy_1 dy_2,$$

which implies that  $\text{Prob}(\xi_1 < c_1, \xi_2 < c_2)$  is an increasing function with respect to  $\rho_{12}$ . This completes the proof.

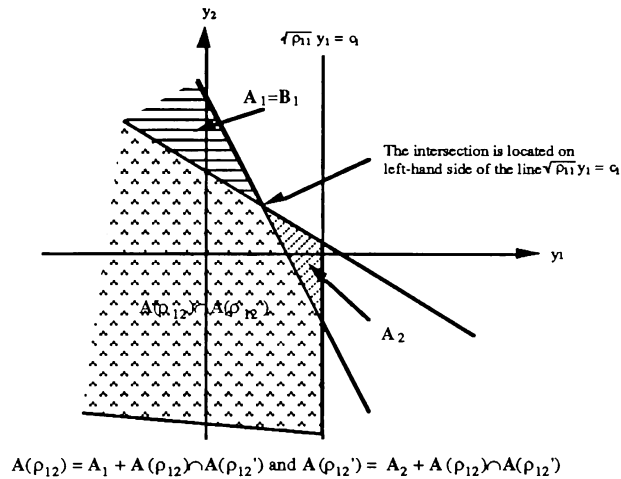


Figure 4a. Definition of  $A_1$  and  $A_2$  for the case  $A_1=B_1$

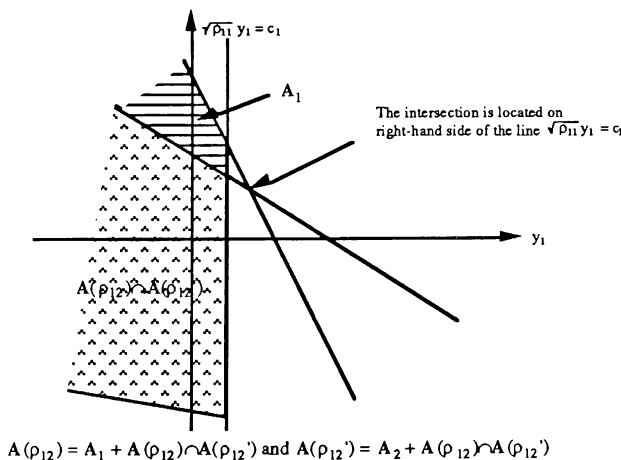


Figure 4b. Definition of  $A_1$  and  $A_2$  for the case  $A_2=0$   
QED

The above theoretical results for  $N=2$  and  $3$  show that in general the larger the correlation between estimation noises the better chance we have to find actual “good” designs. Of course, for high dimension cases our calculation becomes extremely difficult and tedious if not impossible; hence one has to resort experimental methods as we did in the previous section.

## 5. CONCLUSION

We are thus moved to conclude that

*Correlation in estimation error in general can only help the ordinal optimization process*

with the caveat that estimation errors are not biased or correlated in perverse ways.

## ACKNOWLEDGEMENTS

The work in this paper is supported by the Office of Naval Research Grants N00014-90-J-1093 and N00014-89-J-1023, by National Science Foundation Grants ECS-9102346 and CDR-8803012, by Army Grant DAAL-03-86-K-0171, and by Army Research Office Grant DAAL-03-91-G-0194.

## APPENDIX A: UNIFORM NOISE

In the main body of this paper, we have mainly focused on the model in which  $v_1, \dots, v_N$  are i.i.d. normal random variables. As we have argued in this Section 2, this is a reasonable model for most cases of practical interest. Nevertheless, we have also investigated the case in which  $v_1, \dots, v_N$  are i.i.d. uniform random variables to see how our observations can be affected. As it turns out similar conclusions to the normal case can be made in the uniform case as well. In this appendix we present both experimental and analytical results we obtained. Let us first present our experimental results.

Similar to what we did in Section 2, we tested a case in which  $w_1 = v_1$ , and  $w_{n+1} = aw_n + bv_{n+1}$  ( $n = 1, \dots, N-1$ ), where  $v_1, \dots, v_N$  are i.i.d.  $U(-50, 50)$  and  $a$  and  $b$  are two constants such that  $a^2+b^2=1$ . The results are given in Figure 5.

From the figure we can see that

- For the flat case, when the correlated factor “ $|\alpha|$ ” increases from 0 to 1, the number of good designs in the top-ten first decreases slightly and then increases. This implies that the correlation is not always helpful. The reason why this occurs is as follow. We keep the mean and variance of  $w_i$  fixed. But the distribution of  $w_i$  is not  $U(-50, 50)$ . We know that the distribution of the sum of two uniformly distributed variables is not uniform. So we can not get uniformly distributed variables  $w_i$   $i=1, \dots, 200$ , though  $v_i$  and  $w_1$  have uniform distributions. Hence, varying “ $a$ ” causes much more complicated changes in its effect on uniform noise correlated model.

- However, the decrease of the average number of good designs,  $m$ , disappear for  $k>0.5$  corresponding to the linear and steep case. Thus, over all we can still say that **correlation in the estimation noise seldom hurts but generally helps ordinal optimization.**

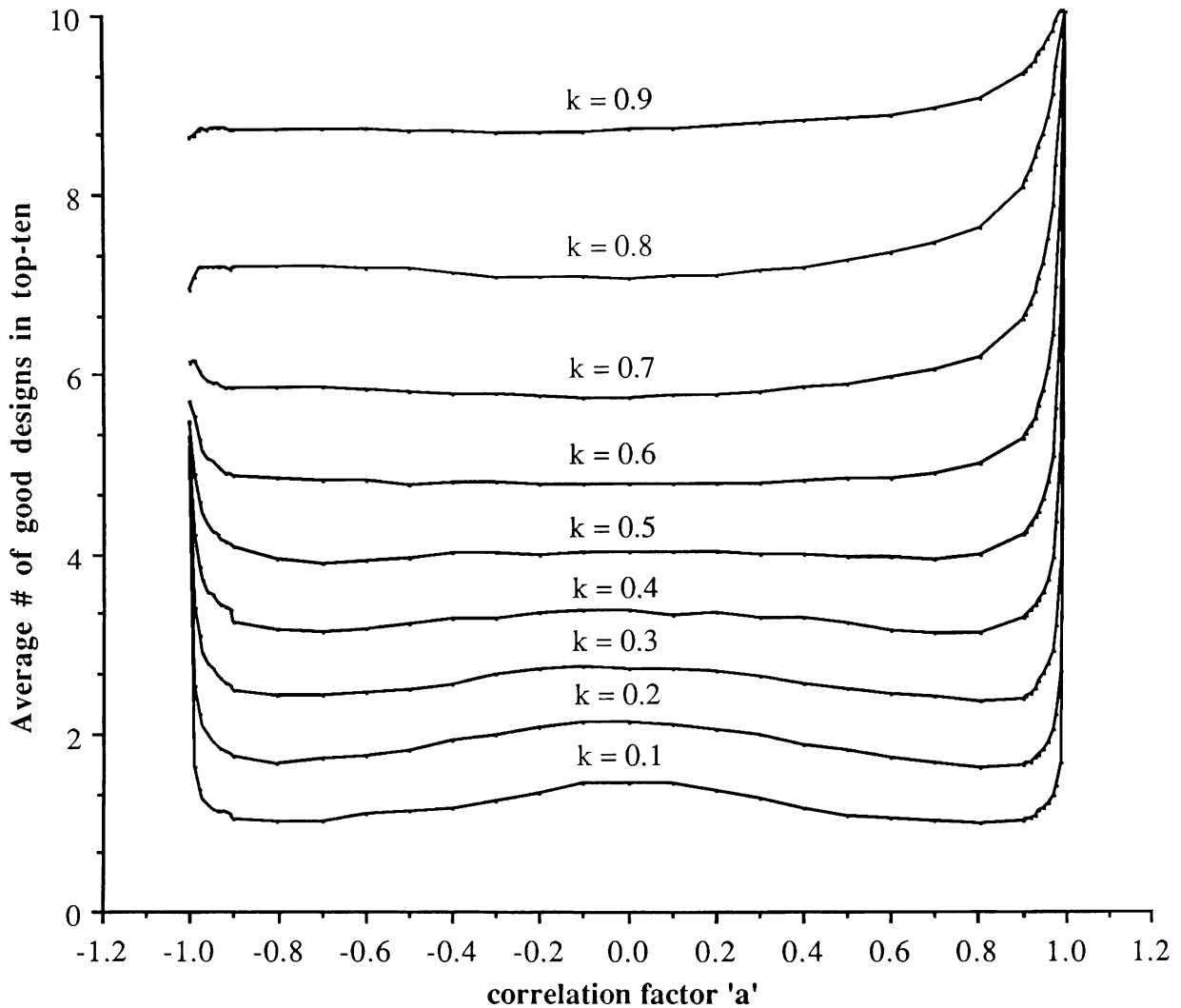


Figure 5: Experimental Results for Uniform Case

Next we investigate the uniform case analytically. Let us again consider the simplest case  $N=2$ , i.e.,  $w_1 = v_1$  and  $w_2 = av_1 + bv_2$ , where we assume that  $v_1$  and  $v_2$  are two independent uniform variables over  $[-1, 1]$ . We will restrict ourselves to  $a>0$  and  $b>0$ . Again, let us calculate  $\text{Prob}(J_{[1]} < J_{[2]})$ . We want to show it is possible  $\text{Prob}(J_{[1]} < J_{[2]})$  be a decreasing function of 'a' (the correlation between  $w_1$  and  $w_2$ ), i.e.,

$$\frac{d\text{Prob}(J_{[1]} < J_{[2]})}{da} < 0.$$

We assume  $d > 1$  and  $(d-1)$  and 'a' are small enough such that

$$-1 \leq \sqrt{\frac{1+a}{1-a}} - \frac{d}{1-a} \leq 1$$

(Recall  $d = J_2 - J_1 > 0$ .)

Note  $b = \sqrt{1-a^2}$ , we have

$$\text{Prob}(J_{[1]} < J_{[2]}) = \text{Prob}(w_1 - w_2 < d)$$

$$= \frac{1}{2} \int_{-1}^1 P\left(v_1 < \sqrt{\frac{1+a}{1-a}} x_2 + \frac{d}{1-a}\right) dx_2$$

$$= \frac{1}{4} \left[ 3 - \frac{1}{\sqrt{1-a^2}} + d \left( \frac{1}{\sqrt{1-a^2}} + \frac{1}{1-a} \right) \right] + \frac{1}{4} d^2 \left( \frac{1}{2(1-a^2)} - \frac{1}{(1-a)\sqrt{1-a^2}} \right)$$

Hence,

$$\left. \frac{d}{da} \text{Prob}(J_{[1]} < J_{[2]}) \right|_{a=0} = \frac{d-d^2}{4} < 0,$$

which implies that  $\text{Prob}(J_{[1]} < J_{[2]})$  is a decreasing function with respect to 'a' when 'a' is very small.

We also note that if d is large enough such that

$$\sqrt{\frac{1+a}{1-a}} - \frac{d}{1-a} \leq -1$$

then  $\text{Prob}(J_{[1]} < J_{[2]}) = 1$  is independent of 'a'. Since large 'a' corresponds to the steep case while small 'a' corresponds to flat case, our results here are consistent with the above experimental results.

**APPENDIX B: NOISE MODEL FOR SIMPLE QUEUEING SYSTEMS**

Example 1. We consider an M/M/1 queue with mean interarrival time  $l=100$ . The objective function is

$$J = \text{mean system time} + s / \text{service time}$$

where s is a constant. The parameter of interest is the service time b. 36 different values of b are chosen which are uniformly distributed over [19.4, 96.4]. We run 200 replications to obtain the correlation between each value of the parameter and a fixed value  $b = 41.4$ . In each replication the run length is 1000 customers. Figure 6a shows the correlation between the values larger than 41.4 and the fixed value 41.4, while Figure 6b shows the correlation between the values smaller than 41.4 and the fixed value 41.4.

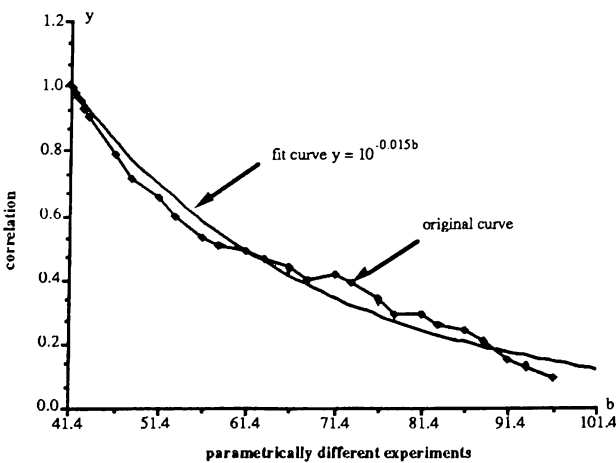


Figure 6a: Correlation between the service time b larger than 41.4 and the fixed value 41.4

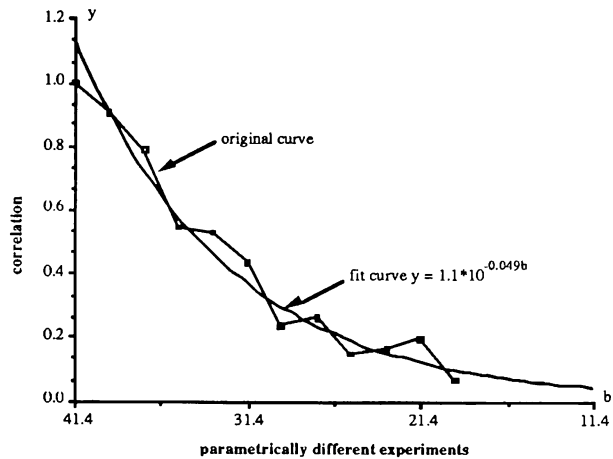


Figure 6b: Correlation between the service time b smaller than 41.4 and the fixed value 41.4

*This shows that our model  $w_{n+1} = aw_n + bv_{n+1}$  is good because the correlation between the estimation errors of two values of the parameter is decreasing exponentially with respect to the difference of the parameters.*

Example 2. This is a two-node Jackson queueing network (see Figure 7).

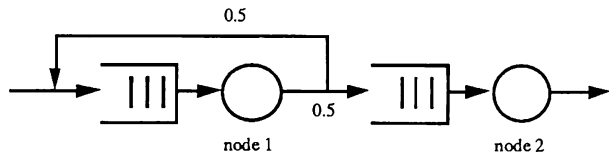


Figure 7: Example 2.

Customers arrive at node 1 with rate  $\lambda_1=0.5$ . Upon service completion at node 1, customers go to node 1 with probability 0.5 and to node 2 with probability 0.5.

The objective function J is the average sojourn time of a customer in the network. We want to minimize under the constraint  $\mu_1 + \mu_2 = 4$ , where  $\mu_i$  is the service rate at node i. 24 different values of  $\mu_1$  are chosen which are uniformly distributed over [1.16, 3.46] and  $\mu_1 = 2.46$  is (arbitrarily) chosen as the reference value. Again we run 200 replications to obtain the correlation between each value of the parameter and the reference value. In each replication the run length is 1000 customers. Figure 8a shows the correlation between the values larger than 2.46 and 2.46 and Figure 8b shows the correlation between the values smaller than 2.46 and 2.46.

*In this case the correlation between the estimation errors of two values of the parameter is decreasing polynomially with respect to the difference of the parameters. So we can not use the simple model  $w_{n+1} = aw_n + bv_{n+1}$  here.*



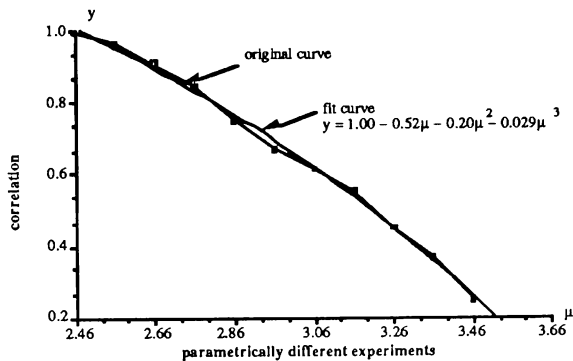


Figure 8a: Correlation between the service rate  $\mu_1$  larger than 2.46 and 2.46

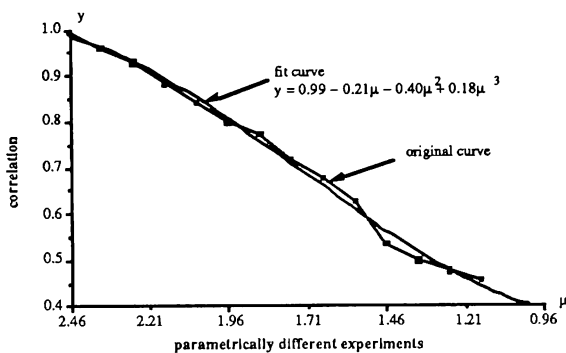


Figure 8b: Correlation between the service rate  $\mu_1$  smaller than 2.46 and 2.46

Based on the above two examples, we draw the following conclusion:

*The correlation between the estimation errors of two values of the parameter is positive and decreasing with respect to the difference of the two values of the parameter. Furthermore if the correlation is decreasing exponentially it can be characterized by the simple model  $w_{n+1} = aw_n + bv_{n+1}$ .*

## REFERENCES

- Y.C.Ho, R.S.Sreenivas, P.Vakili, "Ordinal Optimization of DEDS," *Journal of Discrete Event Dynamic Systems*, 2, #2, 61-88, 1992.
- Vakili, P. Mollamustafaoglu, L.,Ho, Y.C., "Massively Parallel Simulation of a Class of Discrete Event Systems", *Proc. of the IEEE Symposium on the Frontier of Massively Parallel Computation*, 1992.

## AUTHOR BIOGRAPHIES

**MEI DENG** is a Ph.D. student in Division of Applied Sciences at Harvard University. Her research interests include simulation, optimization, and control of discrete event systems and queueing network with applications to communication networks.

**YU-CHI HO** received his S.B. and S.M. degrees in Electrical Engineering from M.I.T. and his Ph.D. in Applied Mathematics from Harvard University. Except for three years of full time industrial work he has been on the Harvard faculty where he is the T. Jefferson Coolidge Chair in Applied Mathematics and Gordon McKay Professor of Engineering..

He has published over 100 articles and three books, and is the recipient of various fellowships and awards including the Guggenheim (1970) and the IEEE Field Award for Control Engineering and Science (1989). He is a fellow of IEEE, a Distinguished Member of the Control Systems Society, and a member of the U.S. National Academy of Engineering.

**JIAN-QIANG HU** is an Assistant Professor in the department of Manufacturing Engineering at Boston University. His research interests include sensitivity analysis, simulation, optimization, and control of discrete event systems and queueing network with applications to manufacturing systems, inventory systems, and communication networks.