Factorial Design for Efficient Experimentation

GENERATING INFORMATIVE DATA FOR SYSTEM IDENTIFICATION

JAMES C. SPALL

Dyson, who hand-built 5,127 prototypes of his vacuum, says, "You have to change one thing at a time. If you make several changes simultaneously, how do you know which has improved the object and which hasn't?"

—Inventor and entrepreneur James Dyson, as quoted in Hemisphere Magazine, the in-flight magazine of United Airlines, November 2005

While James Dyson's eponymous company, Dyson Ltd., based in Malmesbury, United Kingdom, is known to make high-quality vacuum cleaners, Dyson might have saved himself and his company a lot of time and money if he had been aware of the factorial design approach to experimentation. In fact, as discussed below, the statement of Dyson on cause and effect is incorrect. It is possible to learn what improves a system by changing more than one input variable at a time. Furthermore, this learning can be done more efficiently and with the additional benefit of acquiring information about input-variable interactions that cannot be revealed in the one-at-a-time changes advocated by Dyson. Unfortunately, he is not alone in his beliefs. Many in industry, government, and elsewhere share such

NASA/JOHNS HOPKINS UNIVERSITY APPLIED PHYSICS LABORATORY/CARNEGIE INSTITUTION OF WASHINGTON

Digital Object Identifier 10.1109/MCS.2010.937677

1066-033X/10/$26.00©2010IEEE

misconceptions. The aim of this article is to present some of the key ideas in factorial designs and demonstrate how this approach offers both greater efficiency and insight than one-at-a-time changes.

Experimental design, including factorial design and other approaches, provides a structure for determining the values of input variables in engineering and scientific experiments with the aim of extracting as much information as possible relative to the constraints and goals of the analysis. Factorial design is a foundational strategy for experimental design. Although there are many specific implementations of factorial design, all such implementations involve the simultaneous variation of multiple components within the input vector. Factorial designs, therefore, differ fundamentally from classical strategies based on changes in only one input variable at a time.

Factorial design is not a new approach. The statistician R.A. Fisher developed many of the underlying ideas in the 1920s for applications in agriculture. Factorial design has been widely used in process control and related industries since at least the 1950s. Nevertheless, misconceptions remain, as in the Dyson quote given earlier. In particular, while changing one variable in an isolated test is valuable for understanding system sensitivity, simultaneous changes of a factorial design are more useful than a set of one-at-a-time changes when conducting an experiment that encompasses multiple tests.

Consider an input–output process relating the input vector $\boldsymbol{x}$ to the scalar output z, where the dimension of $\boldsymbol{x}$ is $m \geq 1$. In the language of experimental design, $\boldsymbol{x}$ is composed of factors, with each factor representing one of the inputs that the analyst can adjust in the experiment. Factors may be quantitative (often varying continuously) or qualitative (often assigned a discrete scale), with quantitative factors naturally associated with a numerical scale and qualitative factors associated with a labeling scheme or the absence or presence of some system characteristic. Examples of quantitative factors are pressure, temperature, and time in a physical system or money supply and interest rates in a macroeconomic system. Examples of qualitative factors are operation mode in a system, the choice of open- or closed-loop control, or the absence or presence of a human operator, with the relevant components in $\boldsymbol{x}$ being discrete (for example, categorical) to indicate which option is being used. While the components of $\boldsymbol{x}$ may be either continuous or discrete, some statistical approaches, such as the response surface methodology based on gradient descent, restrict the components of $\boldsymbol{x}$ to being continuous; see the section "Response Surface Methodology."

We assume that there is full control over the values, that is, levels, of the factors in producing a set of data. Suppose that the aim is to understand the relationship between the vector $\boldsymbol{x}$ and the output z, including, for example, the *main effect* for each factor, representing the average change in output that follows from a change in the level of each component of $\boldsymbol{x}$. Such information typically governs the design of a control law if the model is to be used as part of a control system. The choice of input levels relates to key issues in system identification, such as enhanced estimation of model parameters, model checking using validation and verification procedures, and mitigation of nuisance terms that are not of direct interest but that affect system performance.

The focus here is the standard 2^m full-factorial design ([1, Chaps. 5–7], [2, Ch. 11], [3, Chaps. 4–6]), which relies on collecting data from experiments having m factors that can be simultaneously varied. We henceforth refer to this design as simply the *2^m factorial design* to distinguish it from other types of factorial designs. As suggested by the presence of the "2" in 2^m, the 2^m factorial design is based on performing runs with each input factor at one of two levels, low and high. Although 2^m factorial designs and other types of factorial designs can be implemented for general input–output relationships, most of the justification for factorials and most of the applications to parameter estimation and hypothesis testing rely on system models that are linear in the underlying parameters $\boldsymbol{\theta}$ that describe the process. The input factors appearing in $\boldsymbol{x}$, however, may enter nonlinearly in these linear models, including possibly through factor interactions. A model that is linear in $\boldsymbol{\theta}$ and possibly nonlinear in $\boldsymbol{x}$ is called a *curvilinear model* [4, Sec. 3.1.1].

The 2^m factorial design is often used in the early stages of experimentation for input screening. In particular, using data collected from the 2^m factorial experiment together with appropriate statistical tests, we can determine whether the output depends in a significant way on the values of specific inputs or functions of inputs. Such screening helps refine the model form, allowing for more precise experimental design and parameter estimation in subsequent experimentation.

Let $\boldsymbol{x}_k$ and z_k denote the input and output for the kth run among the multiple runs of an experiment; we drop the subscript k, writing $\boldsymbol{x}$ and z, when making generic references to inputs and outputs. Thus, for example, if $m = 3$ and two- and three-way interactions are assumed to occur among the input factors, the kth output in the experiment is modeled according to

$$z_k = \beta_0 + \beta_1 x_{k1} + \beta_2 x_{k2} + \beta_3 x_{k3} + \beta_4 x_{k1} x_{k2} + \beta_5 x_{k1} x_{k3} + \beta_6 x_{k2} x_{k3} + \beta_7 x_{k1} x_{k2} x_{k3} + v_k, \quad (1)$$

where $\boldsymbol{\theta} = [\beta_0, \beta_1, \ldots, \beta_7]^T$ represents the vector of unknown parameters, T denotes transpose, x_{ki} represents the ith term in the input vector $\boldsymbol{x}_k$, and v_k represents a random error. As discussed in [2, Sec. 6.3.3], the error term v_k may have various parts, such as experimental and observational errors and constituents therein, but we do not consider this decomposition here. We assume the noise has zero mean unless noted otherwise, that is $E(v_k) = 0$ for all k. The model (1) is nonlinear in the input domain in the

Factorial design is one of the most common and powerful strategies for choosing the values of inputs in experiments.

sense that the mean output $E(z_k)$ depends on the values of $\boldsymbol{x}_k$ lying in a twisted hyperplane in $\mathbb{R}^m$, which is a surface with curved contour lines of constant response $E(z_k)$. Note that 2^m factorial designs are not appropriate for curvilinear models that include more general nonlinearities than interaction terms of the type shown in (1). In particular, 2^m factorial designs are not appropriate for models with squared terms , such as x_{ki}^2, or other nonlinear functions of the components of $\boldsymbol{x}_k$. As discussed in the section "Statistical Analysis, Extensions, and Connections to Stochastic Optimization," alternative factorial design strategies are recommend when such nonlinear terms are significant.

Although formal methods of experimental design, including various forms of factorial design, are commonly used to choose input levels for physical processes, these methods are also useful when selecting input levels for computer simulation models of processes [5]. Typical uses for the data coming out of a simulation are sensitivity analysis or the construction of a design model in the form of a metamodel or response surface representing a simplified form of the full simulation. Design models are useful in building controllers for systems because the analytical analysis of open- and closed-loop system behavior with a design model is less complicated than with the full simulation [6, pp. 20–21]. One category of systems for which simulation models are often used is continuous-time systems that have a change of state at random points of time; the design and control of such *discrete-event dynamic systems* is an area of significant interest in the field of control [4, Sec. 5.3]. As with physical processes, factorial design is used to increase the value of the simulation runs, relative to simulation runs with nonfactorial levels for the inputs, toward understanding the system under study.

There are some differences in the implementation of experimental design with simulations and the implementation with physical systems. For example, m is often much larger with simulations than it is with physical systems. Having m on the order of 100 or 1000 is not unusual when generating data from simulations, while $m \leq 10$ is typical when working with physical data [7, Sec. 2.2]. Another difference is that there is greater control over the randomness in simulations through control of the random number seed and, perhaps, through the design of the underlying pseudorandom number generators. This control of randomness allows for the implementation of the method of common random numbers [4, Sec. 14.4], [5, pp. 93–94], [8, Sec. 11.2] or alternative variance reduction techniques [8, Ch. 11], thereby enhancing the value of the experiments in the intended application.

Factorial designs balance multiple objectives by providing an informative distribution of inputs throughout the domain of interest, being reasonably simple to implement, allowing the experiment to be built up sequentially (useful if subsequent data become available or if it is desirable to add terms to the model), being efficient in terms of not requiring an excessive number of data points (which may require fractional factorial designs in large-dimensional problems [1, Ch. 8]), and providing accurate parameter estimates from the given budget of n runs. Nevertheless, factorial designs are not necessarily optimal in any formal sense. Thus, factorial designs are not as widely discussed in the control literature as the *optimal design* class of methods within experimental design [9], where $\boldsymbol{x}$ is chosen to optimize an objective function. A discussion of the relationship between factorial design and optimal design is given in the section "Connections to Regression, Orthogonality, and Optimal Design."

The lack of attention to factorial designs in the control literature is at odds with the extent of use in control applications. Various implementations of factorial designs are widely used in industrial control and elsewhere for the historical and technical reasons given above. Although many industrial implementations of factorial design are not discussed in the public domain, there are at least a few applications covered in the control literature, including proportional-integral-derivative (PID) controller design [10], neural networks for supervisory control [11], and modeling for automotive engine design and optimization [12].

FUNDAMENTALS OF FACTORIAL DESIGN

The one-at-a-time approach to choosing input levels involves making small changes to each input while holding the remaining inputs constant. While this approach can be used to construct finite-difference approximations to gradient vectors, it is usually an inefficient way to learn about the effect of the input factors on the output and does not provide information about the interactions of the input factors. Factorial design overcomes the shortcomings of one-at-a-time changes by increasing the utility of a given number of experimental runs in terms of forming more accurate estimates of the parameters associated with the main effects in $\boldsymbol{\theta}$, corresponding to β_0, β_1, β_2, and β_3 in (1), and by revealing information about interactions that would not be available when changing only one input at a time, corresponding to β_4, β_5, β_6, and β_7 in (1).

The 2^m factorial design is appropriate when the number of inputs m is not too large. Further, the 2^m factorial design provides the foundation for related strategies, such as fractional factorial [1, Ch. 8], when m is larger. With quantitative factors, the low or high interpretation is natural; with qualitative factors, the low or high value might indicate the absence or presence of an entity. For example, in a feedback system, low can denote no closed-loop control while high can denote closed-loop control. Let "$-$" represent a low value of the given input and "$+$" a high value, where "low" and "high" are relative to the definition of the given component of the $\boldsymbol{x}$ vector. The unnormalized numerical values corresponding to a $+/-$ symbol for input 1 (say, x_{k1}) and for input 2 (x_{k2}) may differ significantly according to the physical interpretations and units of measurement for the various input factors. It is often convenient, however, to normalize the numerical values to ± 1.

Many implementations of factorial design rely on *replications*, where each replicate represents a set of 2^m runs. The replicates are mutually independent. Multiple replicates are used to decrease the variance of the outputs at different input levels and, correspondingly, to increase the ability to detect differences in the influence of input factors. For example, in carrying out analysis of variance (ANOVA) statistical tests on models such as (1), replications can be used to sharpen the statistical inference and increase the precision of the associated estimates that indicate whether one factor is more influential than another. Note, however, that it is assumed in ANOVA that the noise terms v_k are independent and identically distributed (i.i.d.) across all runs in an experiment. That is, the noise is i.i.d. across all runs both within a replicate and across all replicates. In practical settings where the noises for the runs inside a replicate are not necessarily i.i.d., replications can be used with non-ANOVA methods, such as simple averaging, to analyze the influence of individual input factors, or interactions of two or more input factors, on the output. See the section "Statistical Analysis, Extensions, and Connections to Stochastic Optimization."

Let us consider the case $m = 3$ to motivate the key ideas behind 2^m factorial designs. In this case, one replicate of the design produces $2^3 = 8$ output values. Table 1 shows the eight possible input combinations for one experiment, while Figure 1 depicts this design as a cube diagram. With replications, each output in a table such as Table 1 is the sample mean of the outputs for the multiple replicates. Thus, for example, an experiment of ten replicates of a 2^3 factorial design produces 80 observations of z. The representation in Table 1 is the *standard (Yates) order* for the input combinations in a factorial design [1, p. 227].

In collecting data from physical systems, it is usually recommended to randomize the order of the runs or conduct the runs in homogenous groups, called *blocks*, to help cope with environmental or other nuisance effects that can introduce systematic unmodeled biases in the data. In system identification applications, nuisance effects may manifest themselves in uncertain model parameters that are not of primary interest [13]. For example, a slow increase in ambient temperature in a laboratory might affect the results in a biological experiment or the improving performance of a human operator over time might affect the quality of the output from a machine in a factory experiment. Randomization helps mitigate these unmodeled effects. In the context of replications of a 2^m factorial experiment, randomization might be implemented by carrying out the runs in each replicate in a random order.

Blocking is a related strategy used to control exogenous effects that are not the focus of the study. The runs within a block are expected to be more alike than runs from a mixture of inside-block and outside-block settings. For example, all runs might be conducted on a particular day to help remove the effect of day-to-day weather variations. The randomized complete block design ([1, Ch. 4] or [2, Sec. 9.2])

TABLE 1 Input and output combinations for a 2^3 factorial design with $\boldsymbol{x}_k = [x_{k1}, x_{k2}, x_{k3}]^T$. With multiple replicates, the indicated output in the rightmost column represents the sample mean of observations from multiple independent runs at the indicated input levels. Randomization of the order of the runs is usually recommended with multiple replicates for physical systems.

Design Point $\boldsymbol{x}_k$	Input 1 x_{k1}	Input 2 x_{k2}	Input 3 x_{k3}	Output
$\boldsymbol{x}_1$	$-$	$-$	$-$	z_1
$\boldsymbol{x}_2$	$+$	$-$	$-$	z_2
$\boldsymbol{x}_3$	$-$	$+$	$-$	z_3
$\boldsymbol{x}_4$	$+$	$+$	$-$	z_4
$\boldsymbol{x}_5$	$-$	$-$	$+$	z_5
$\boldsymbol{x}_6$	$+$	$-$	$+$	z_6
$\boldsymbol{x}_7$	$-$	$+$	$+$	z_7
$\boldsymbol{x}_8$	$+$	$+$	$+$	z_8

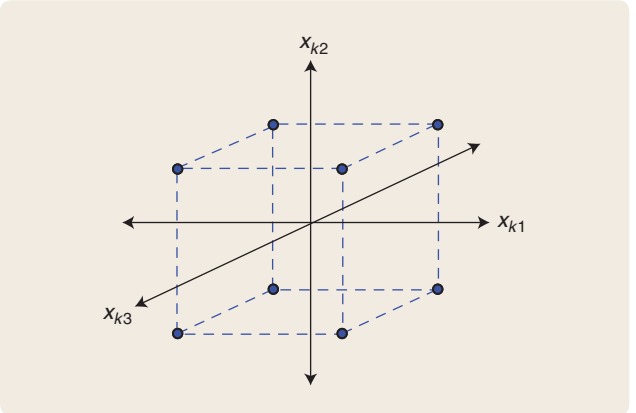

FIGURE 1 Graphical depiction of the input points for a 2^3 factorial design. The indicated points are symmetric with respect to the origin. In contrast to one-at-a-time changes, no input point lies on an axis.

uses randomization within blocks to further reduce nuisance effects. Alternative blocking strategies for randomization can apply with one or more replicates [1, Chaps. 4 and 7]. When a computer simulation, rather than a physical system, provides the experimental data, there are typically no hidden effects that need to be mitigated. Thus, randomization and blocking are usually not necessary with simulation-based data [5, pp. 52–53].

MAIN EFFECTS AND INTERACTION EFFECTS

We now consider direct methods by which experimental design can be used to help determine relationships between inputs and outputs. The related issue of estimation for model parameters $\boldsymbol{\theta}$, such as the eight values β_i in (1), is discussed in the sections "Connections to Regression, Orthogonality, and Optimal Design" and "Response Surface Methodology." Statistical inference based on the direct methods can be made under weaker non-i.i.d. assumptions on the noise v_k than the i.i.d. assumptions typically needed with regression-based methods for estimating $\boldsymbol{\theta}$. Non-i.i.d. noise arises frequently, for example, in applications where computer simulations produce the outputs z_k.

Continuing with the example of Table 1, we can use the information represented in the table to determine the main effects and interaction effects for the three inputs. Let effect_i denote the main effect for input i. For each main effect, we are interested in the average change in the output as a given input is changed from the low $(-)$ value to the high $(+)$ value. Hence, for input 1, the average increase in output for a change in x_{k1} is

$$\text{effect}_1 = \frac{(z_2 - z_1) + (z_4 - z_3) + (z_6 - z_5) + (z_8 - z_7)}{4}. \quad (2)$$

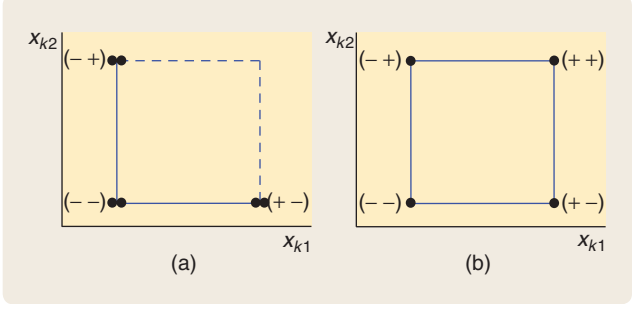

FIGURE 2 Experimental strategies to obtain equivalent information for estimating each main effect. The plots assume $m = 2$, where each dot represents one run. Both plots (a) and (b) show two relevant differences in output for each main effect. Plot (a) depicts the one-at-a-time approach, with two runs at each of three input points for a total of six observations. Plot (b) depicts the 2^2 factorial approach with one run at each of four distinct input points. Hence, to yield equivalent information, one-at-a-time experimentation has a cost in number of runs that is 6/4 times the cost of factorial design.

Note that (2) can equivalently be written as the difference between the mean response at the + values and the mean response at the − values, that is,

$$\text{effect}_1 = \frac{(z_2 + z_4 + z_6 + z_8) - (z_1 + z_3 + z_5 + z_7)}{4}.$$

Likewise, for the remaining main effects for $m = 3$, we have

$$\text{effect}_2 = \frac{(z_3 + z_4 + z_7 + z_8) - (z_1 + z_2 + z_5 + z_6)}{4},$$
$$\text{effect}_3 = \frac{(z_5 + z_6 + z_7 + z_8) - (z_1 + z_2 + z_3 + z_4)}{4}.$$

The ideas above apply in a straightforward way to the general m-input problem, where the main effect for input i is

$$\text{effect}_i = \frac{\sum_{k \text{ with "+" factor } i} z_k - \sum_{k \text{ with "−" factor } i} z_k}{2^{m-1}}. \quad (3)$$

Note that the denominator 2^{m-1} reflects the fact that (3) can be interpreted as the average of 2^{m-1} differences, a generalization of the denominator in (2) for the setting $m = 3$.

We now apply the ideas above to illustrate the efficiency of factorial designs for estimating the main effects relative to one-at-a-time changes. Consider $m = 2$, as illustrated in Figure 2. For one-at-a-time experimentation, the effect of a change in x_{k1} is determined by the difference between observations at the points $(--)$ and $(+-)$, while the effect of a change in x_{k2} is determined by the difference at the points $(--)$ and $(-+)$. In contrast, with factorial design, the change in system output due to a change in x_{k1} is measured by the difference between observations at $(--)$ and $(+-)$ as well as by the difference at $(-+)$ and $(++)$. Likewise, the change in system output due to a change in x_{k2} is measured by the difference between observations at $(--)$ and $(-+)$ as well as by the difference at $(+-)$ and $(++)$. With one-at-a-time changes it is necessary to collect observations from two runs at each input to provide the same level of information, that is, two differences, available from the four factorial runs. Figure 2(a) depicts this requirement for two runs at each point. Hence, the ratio of the required one-at-a-time observations over the required factorial observations to obtain equivalent information is 6/4. In addition, as discussed below, if the input variables interact significantly, then the one-at-a-time method can give misleading indications of the effects of a change, while the factorial method can reveal the interaction effects.

Let us extend the efficiency calculation for $m = 2$ associated with Figure 2 to the case $m \geq 3$. For factorial experiments, there are 2^{m-1} differences between observations available for each factor. In contrast, for one-at-a-time experiments, $(m + 1)2^{m-1}$ runs are required to obtain 2^{m-1} differences between observations for each factor. The $m + 1$ term comes from the need to obtain 2^{m-1} observations at the nominal "low" value for all factors simultaneously plus

the need to obtain 2^{m-1} observations for each one-at-a-time change of all of the m factors. Hence, the ratio of the number of observations needed in a one-at-a-time experiment to the number required in a factorial experiment grows linearly according to the expression

$$\frac{(m+1)2^{m-1}}{2^m} = \frac{m+1}{2}. \tag{4}$$

Figure 3 depicts this expression for several values of m, showing, for example, that a factorial experiment requires only 1/8th the number of observations as a one-at-a-time experiment to achieve the same accuracy in estimating the main effects at $m = 15$.

An alternative derivation of relative efficiency is in terms of the variances of the main effects. Let the noise terms v_k be i.i.d. with $\sigma^2 = \text{var}(v_k)$. From (3), the variance of a main effect estimate with one set of 2^m runs for the factorial approach is $2^m\sigma^2/(2^{m-1})^2 = \sigma^2/2^{m-2}$. In contrast, the variance of a main effect estimate for one set of $m+1$ runs in the one-at-a-time method is $2\sigma^2$. Hence, the sample mean of 2^{m-1} independent replicates of the one-at-a-time method has the same variance as one replicate of the factorial method. Therefore, the ratio of the number of runs needed in one-at-a-time relative to factorial to achieve the same variance is $(m+1)2^{m-1}/2^m = (m+1)/2$, as in (4).

The main effects calculations are useful for providing the average response to changes in the individual input factors. These calculations, however, do not take into account the fact that the change in response due to a change in, say, input 1 might depend on the level of input 2. If so, the inputs are said to interact. For $m = 2$, Figure 4 gives an example of the difference between the mean responses with and without interaction. Figure 4(a), depicting no interaction, shows that the change in $E(z_k)$ with a change in x_{k2} does not depend on x_{k1}; Figure 4(b), depicting interaction, shows that the change depends on x_{k1}. It is assumed that $E(v_k)$ is not a function of x_k in Figure 4; in fact, it is almost always assumed in practice that $E(v_k) = 0$. There are numerous ways in which variables can interact beside the interaction shown in Figure 4, including higher-way interactions when $m > 2$, nonlinear interactions different than the type in (1), and block-factor level (called block-treatment) interactions for randomized complete block design discussed above; see, for example, the case study in [2, Sec. 11.5].

Factorial design provides a method for measuring interactions. Consider two indices $i_1, i_2 \in \{1, 2, \ldots, m\}$. Then the two-way interaction effect between inputs i_1 and i_2, denoted $\text{effect}_{i_1 i_2}$, is measured according to half of the difference between the effect of i_1 when input i_2 is held at its + value and the effect of i_1 when input i_2 is held at its − value. These calculations are symmetric in the sense that the interaction between inputs i_1 and i_2 is identical to that between inputs

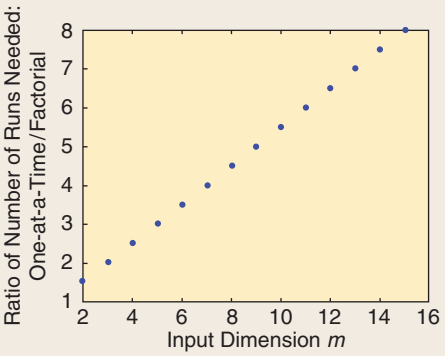

FIGURE 3 Relative efficiency of one-at-a-time and factorial design. The plot compares the number of runs needed to obtain the same number of differences in output to measure the main effects. Equivalently, the plot shows the relative number of runs needed to achieve the same variance of a main effects estimate under the assumption of independent and identically distributed noise variance. The indicated linear relationship also applies for larger m.

i_2 and i_1, that is, $\text{effect}_{i_1 i_2} = \text{effect}_{i_2 i_1}$. For $m = 3$, the interaction effect between inputs 1 and 2 is

$$\text{effect}_{12} = \frac{1}{2}\left[\frac{(z_4 - z_3) + (z_8 - z_7)}{2} - \frac{(z_2 - z_1) + (z_6 - z_5)}{2}\right]. \tag{5}$$

Analogous expressions apply for the interactions between inputs 1 and 3 and inputs 2 and 3.

As in the calculation of the main effects, there is a convenient tabular method for determining interaction effects. The entries in Table 1, or its analogue for problems with $m \neq 3$, correspond to the *input matrix*, also called the *design matrix* or *model matrix*. When considering the interaction between inputs i_2 and i_1, for example, we create an additional column in the input matrix by multiplying the signs from the columns for inputs 1 and 2,

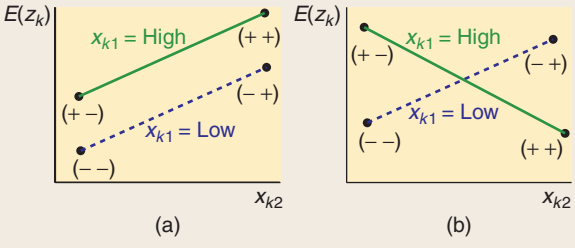

FIGURE 4 Example plots comparing the mean of the output z_k when x_{k1} and x_{k2} do not interact and when they interact. In plot (a), the change in mean output due to a change in x_{k2} is independent of the value of x_{k1}; likewise when interchanging x_{k1} and x_{k2}. When the variables interact, as in plot (b), the level of x_{k1} affects the change in output relative to a change in the level of x_{k2}; likewise when interchanging x_{k1} and x_{k2}. In contrast to one-at-a-time experimentation, factorial design in conjunction with statistical tests can detect interactions.

Factorial Design for a Circuit Board, Part 1: Formulation and Data

A factorial experiment was conducted shortly before the launch of the MErcury Surface, Space ENvironment, GEochemistry, and Ranging (MESSENGER) spacecraft from Cape Canaveral, Florida on August 3, 2004. MESSENGER is scheduled to begin orbiting the planet Mercury on March 11, 2011 after the gravity assists of one flyby of Earth, two flybys of Venus, and three flybys of Mercury. Information on the timeline and scientific goals is at the official mission site (http://messenger.jhuapl.edu/).

The factorial experiment was carried out to determine the cause of an increase in the resistance of the resistors used in circuit boards on the spacecraft, an increase that was observed as the spacecraft sat on the launch pad. The change in resistance threatened to delay the launch unless the cause was found and, as needed, corrective action was taken. The factorial experiment was conducted at the Johns Hopkins University, Applied Physics Laboratory. Engineers identified three factors suspected of contributing to the increase in resistance:

- *Factor 1, solder temperature*. The manufacturer recommends a solder tip temperature of 500 °F when installing resistors on the circuit boards. However, a higher temperature of 700 °F was used in the installation in the board on the spacecraft. The laboratory experiment used 500 °F as the − value and 700 °F as the + value.
- *Factor 2, resistor batch*. The resistors installed on the circuit board in the spacecraft were drawn from both the older and newer of two batches available from the manufacturer, leading to a concern that the batch choice might have affected the change in measured

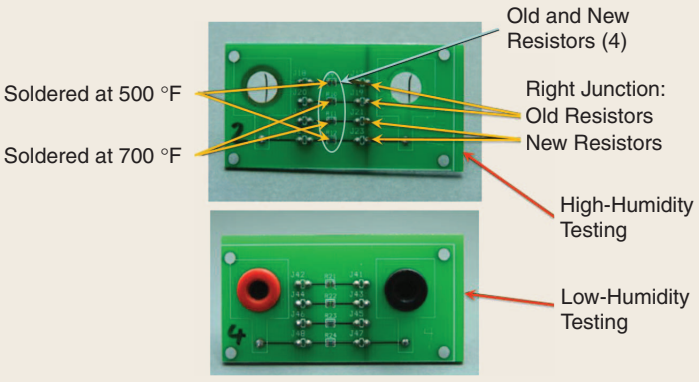

FIGURE S1 Circuit boards used to determine the cause of increased resistance in the MESSENGER spacecraft. The boards are used in a laboratory experiment that is based on one replicate of a 2^3 factorial design. The top board is tested in a steam bath at 100% humidity; the bottom board is tested at ambient humidity levels in the laboratory. The resistor configuration is the same for both boards. The total number of resistors is eight, consistent with the 2^3 factorial design.

where like signs yield + and unlike signs yield −. The interaction effect for a particular pair of inputs is then the weighted average of the outputs, weighted by the sign from the corresponding column in the input matrix. The denominator in the averaging process is 2^{m-1}, as with the main effects above.

TABLE 2 Signs (±) to apply to output values in calculating specified two-way interactions in a 2^3 factorial design. The outputs in the right-most column are combined according to the indicated signs to estimate an interaction at the top of columns 1–3. As in Table 1, when the design is replicated, the outputs in the right-most column represent the sample mean of observations from multiple replicates.

Inputs 1 and 2	Inputs 1 and 3	Inputs 2 and 3	Output
+	+	+	z_1
−	−	+	z_2
−	+	−	z_3
+	−	−	z_4
+	−	−	z_5
−	+	−	z_6
−	−	+	z_7
+	+	+	z_8

As an illustration, Table 2 shows the corresponding three columns for the three possible interactions in the case $m = 3$, built from the individual input values in Table 1. For example, using the first column of the Table 2, effect_{12} can be written as

$$\text{effect}_{12} = \frac{z_1 - z_2 - z_3 + z_4 + z_5 - z_6 - z_7 + z_8}{4},$$

which is identical to (5).

It is possible to consider higher-way interactions, $\text{effect}_{i_1 i_2 i_3}$, $\text{effect}_{i_1 i_2 i_3 i_4}$, up to a maximum of m-way interaction. For example, three-way interactions represent half the difference between the two-way interactions at the two levels of a third factor. If two-way or higher interactions are present, it is not always possible to conclude that a significant main effect is due solely to the change in the given input from − to +. The magnitude and possibly direction of the change in output can be affected by the levels of other inputs in the problem. The analysis of the

TABLE S1 Results of one set of runs in a 2^3 factorial experiment in the problem of increased resistance in the circuit board. Runs are shown in standard order, as in Table 1. The output z is the percentage change in resistance over a 24-h period relative to the initial resistance in the period; a positive sign for z is associated with an increase in resistance, whereas a negative sign is associated with a decrease in resistance.

Solder Temp. (s)	Resistor Batch (r)	Humidity Level (h)	Initial Resistance (Ohms)	Final Resistance (Ohms)	z
−	−	−	90.87	90.87	0
+	−	−	90.89	90.88	−0.01
−	+	−	243.05	243.04	0
+	+	−	242.98	242.96	−0.01
−	−	+	90.68	200.00	120.6
+	−	+	90.71	198.00	118.3
−	+	+	241.64	244.43	1.155
+	+	+	241.63	248.90	3.009

resistance. Each run of the experiment used resistors from either the old (− level) batch or new (+ level) batch.

- *Factor 3, humidity level during testing.* The presence of high humidity during environmental testing of the circuit board on the launch pad was considered a possible cause of the measured increase in resistance. In the experiment, the – value represented nominal (ambient) humidity in the laboratory, while the + value represented 100% humidity as implemented in a steam bath.

Note that solder temperature and humidity level correspond to quantitative factors, while the batch choice is qualitative. The output variable z represents the percentage increase in resistance over a 24-h period. The use of percentage increase removes the effect of the different nominal levels of resistance in the old and new batches. Only a small number of runs was possible in the experiment because the problem of increased resistance was identified only after the spacecraft was mounted on the rocket at Cape Canaveral and the scheduled launch date was imminent. Therefore, only one set of runs in a 2^3 factorial experiment is used, producing eight output values z. Further, one set of runs is sufficient for statistical inference purposes due to the low levels of noise in the process, which follows from the highly accurate measuring apparatus applied to the high-precision resistors.

Figure S1 shows the resistors in the factorial experiment mounted on two separate circuit boards. The eight resistors shown in the two boards represent the eight possible input combinations at levels + or – for each factor. The factorial design and experimental output are shown in Table S1. Conclusions and statistical results on the significance of main and two-way interaction effects are shown in Part 2 of the example in a separate sidebar.

interaction effects provides the basis for determining such information.

An example of the formulation of a 2^m factorial design is given in "Factorial Design for a Circuit Board, Part 1: Formulation and Data." This example pertains to a 2^3 factorial experiment that was carried out prior to the launch of a spacecraft.

CONNECTIONS TO REGRESSION, ORTHOGONALITY, AND OPTIMAL DESIGN

We now describe connections between factorial design and regression modeling, orthogonal matrices, and the optimal design approach to choosing inputs levels. The output data z are used to estimate a p-dimensional model parameter vector $\boldsymbol{\theta}$.

Consider a linear model of the form

$$\boldsymbol{Z}_n \equiv \boldsymbol{H}_n\boldsymbol{\theta} + \boldsymbol{V}_n, \tag{6}$$

where $\boldsymbol{Z}_n = [z_1, z_2, \ldots, z_n]^T$, $\boldsymbol{H}_n$ is an $n \times p$ input matrix with kth row containing the input combinations associated with the relevant main effects and interaction effects for the associated observation z_k, $\boldsymbol{\theta} = [\beta_0, \beta_1, \ldots, \beta_{p-1}]^T$, and $\boldsymbol{V}_n = [v_1, v_2, \ldots, v_n]^T$. When considering model (1), for example, the kth row of H_n has $p = 8$ entries: $1, x_{k1}, x_{k2}, \ldots, x_{k1}x_{k2}x_{k3}$. Statistical tests on components of $\boldsymbol{\theta}$ correspond to tests on the associated main or interaction effect. If $(\boldsymbol{H}_n^T\boldsymbol{H}_n)^{-1}$ exists, then the standard least-squares estimate for $\boldsymbol{\theta}$ is $\hat{\boldsymbol{\theta}}^{(n)} \equiv (\boldsymbol{H}_n^T\boldsymbol{H}_n)^{-1}\boldsymbol{H}_n^T\boldsymbol{Z}_n$, and the associated covariance matrix for $\hat{\boldsymbol{\theta}}^{(n)}$ is

$$\operatorname{cov}(\hat{\boldsymbol{\theta}}^{(n)}) = (\boldsymbol{H}_n^T\boldsymbol{H}_n)^{-1}\operatorname{cov}(\boldsymbol{V}_n)(\boldsymbol{H}_n^T\boldsymbol{H}_n)^{-1}. \tag{7}$$

Linear model (6) is widely used in systems and control applications [9], [14], [15]. Note that the regression estimates for $\beta_1, \ldots, \beta_{p-1}$ are one half of the direct effects estimates given by (3) and (5) because the regression estimate for each β_i measures the output change from a unit change in corresponding input while the direct estimate is based on the two-unit change of −1 to 1 [1, p. 213]. Generalized least squares is sometimes used when the noise terms do not have the same variance and when the matrix $\operatorname{cov}(\boldsymbol{V}_n)$ is known [2, pp. 125–127], [4, p. 71].

Sequential Construction of Orthogonal Input Matrices

An elegant sequential form exists for building an input matrix at input-dimension m from an input matrix at dimension $m-1$. Following [3, pp. 63–65] and [5, p. 34], Figure S2 depicts the sequential relationship as the input dimension m increases, where, per convention, the lowest input (–) is given the value –1 and the highest input (+) is given the value 1. The figure pertains to the special case of a model $z_k = \beta_0 + \beta_1 x_{k1} + \beta_2 x_{k2} + \cdots + \beta_m x_m + v_k$ having only main effects. As in Table 1, the representation in Figure S2 is the standard order for the input rows in a factorial design [1, p. 227]. Alternative row orderings are statistically equivalent but are not as convenient for the sequential representation in the figure. Table 1 is the special case $m = 3$ of Figure S2.

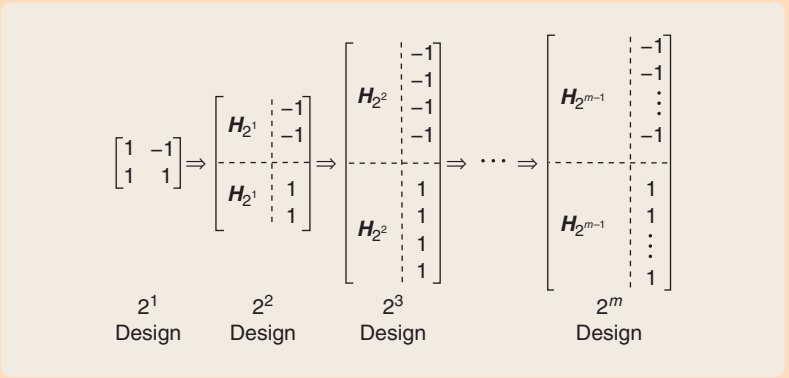

FIGURE S2 Input matrices $\boldsymbol{H}_n$ for factorial designs applied to main effects models with number of inputs 1, 2, 3, …. , m, corresponding to $n = 2^1, 2^2, 2^3, \ldots, 2^m$ measurements. The diagram shows how the input matrix at design 2^m is built from the input matrix at design 2^{m-1} when the rows of the input matrix are in standard order. The sequential construction of input matrices leads to orthogonality of factorial designs at each value of m.

The method of induction can be used with the sequential relationship for constructing input matrices in Figure S2 to show that orthogonality for $\boldsymbol{H}_n$ is true, where the required diagonal matrix takes the form $\boldsymbol{H}_n^T\boldsymbol{H}_n = 2^m\boldsymbol{I}_{m+1}$ for $m \geq 1$ and the number of replicates is $N = 1$. In particular, for the initialization step $m = 1$ with $n = 2^1$, we have

$$\boldsymbol{H}_n^T\boldsymbol{H}_n = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} = 2\boldsymbol{I}_2.$$

Then, for the inductive step from $m-1$ to m, where $n = 2^m$ and $\mathbf{0}$ is an m-dimensional vector, we have

$$\boldsymbol{H}_n^T\boldsymbol{H}_n = \begin{bmatrix} \boldsymbol{H}_{2^{m-1}}^T & \boldsymbol{H}_{2^{m-1}}^T \\ -1,-1,\ldots,-1 & 1,1,\ldots,1 \end{bmatrix}\begin{bmatrix} & -1 \\ \boldsymbol{H}_{2^{m-1}} & -1 \\ & \vdots \\ & -1 \\ & 1 \\ \boldsymbol{H}_{2^{m-1}} & 1 \\ & \vdots \\ & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 2^{m-1}\boldsymbol{I}_m + 2^{m-1}\boldsymbol{I}_m & \mathbf{0} \\ \mathbf{0}^T & 2^m \end{bmatrix} = 2^m\boldsymbol{I}_{m+1}.$$

Therefore, $\boldsymbol{H}_n$ is orthogonal for general m, as desired.

One key property of factorial designs that is also closely related to linear regression is orthogonality. Orthogonality for the input design occurs when the inner product of each pair of distinct columns vectors within $\boldsymbol{H}_n$ is zero. Equivalently, $\boldsymbol{H}_n^T\boldsymbol{H}_n$ must be a diagonal matrix.

The 2^m factorial design, when applied to models such as (1) with main and interaction effects, leads to input matrices that are orthogonal [16, p. 73]. Let $\boldsymbol{I}_p$ denote the $p \times p$ identity matrix and, consistent with standard practice [2, p. 500], [3, pp. 61–62], assume that the inputs are normalized to satisfy $-1 \leq x_{ki} \leq 1$ for all k and i. "Sequential Construction of Orthogonal Input Matrices" considers the special case of models including only main effects. An illustration of orthogonality is apparent by examining the input matrix for one set of runs in the design for $m = 3$ associated with (1) (so $n = 2^3$)

$$\boldsymbol{H}_8 = \begin{bmatrix} 1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 & 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 \\ 1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}. \quad (8)$$

Note that $\boldsymbol{H}_8^T\boldsymbol{H}_8 = 8\boldsymbol{I}_8$ for a single replicate and, more generally, $\boldsymbol{H}_n^T\boldsymbol{H}_n = n\boldsymbol{I}_p$ for $n = N2^m$ and $p \leq 2^m$, where N is the number of replicates. We have $p = 2^m$ when the model includes the constant β_0 and all main and interaction effects up to m-way interactions. We have $p < 2^m$ if the constant β_0 is not included or at least one main or interaction effect is not included.

Although orthogonality for $\boldsymbol{H}_n$ is not required for the validity of the regression estimate, the estimates for all pairs of components of $\boldsymbol{\theta}$ are uncorrelated when the v_k are uncorrelated and $\boldsymbol{H}_n$ is orthogonal. That is, from (7), $\text{cov}(\hat{\beta}_i, \hat{\beta}_j) = 0$ for all $i \neq j$, where $\hat{\beta}_i$ represents the estimate of β_i, corresponding to the $(i+1)$st component in $\boldsymbol{\theta}$. It is, however, often desirable that the components be mutually independent, not just uncorrelated, so that statistical tests can be carried out for each component of $\boldsymbol{\theta}$ separately without having to account for dependence of the estimates.

If it is assumed that the n noise terms v_k are both independent and normally distributed, then each $\hat{\beta}_i$ is formed from a linear combination of independent normally distributed random terms. Hence, mutual independence of the components in $\hat{\boldsymbol{\theta}}^{(n)}$ follows from the uncorrelatedness of the $\hat{\beta}_i$ being equivalent to independence under multivariate normality. Of course, in practice, noise terms are not typically normally distributed, but asymptotic (large n) multivariate normality of $\hat{\boldsymbol{\theta}}^{(n)}$ may follow from central-limit-theorem effects [15, p. 556]. Given such asymptotic normality, the components in $\hat{\boldsymbol{\theta}}^{(n)}$ are approximately independent for sufficiently large n. However, central-limit-theorem effects do not always apply with linear regression, a prominent example being the Kalman filter state estimate, which can be interpreted as a linear-regression estimate [17]. When the state or process noises in a linear state-space model are nonnormally distributed, the central limit theorem may or may not apply to the Kalman filter estimate, depending largely on the stability properties of the state equation [18], [19]. Such counterexamples to asymptotic normality in regression estimation show that care is needed when assigning confidence bounds to the components in $\hat{\boldsymbol{\theta}}^{(n)}$ when the noise terms are not normally distributed.

Alternatives to 2^m factorial designs and curvilinear models of the form (1) can achieve orthogonality or partial orthogonality, the latter corresponding to settings where some—but not all—of the components within $\hat{\boldsymbol{\theta}}^{(n)}$ are uncorrelated. For example, a *star design* [4, pp. 483–484] that places points symmetrically along all axes, which leads to $2m$ runs, is orthogonal. However, as a form of one-at-a-time changes, the star design alone does not offer the advantages of factorial design. One example of partial orthogonality is a 2^m factorial design applied to a quadratic-input polynomial model, as discussed in [4, Ex. 17.8]. Treatments of orthogonality for a variety of factorial designs are provided in [20] and [21, Chaps. 4 and 8].

Orthogonality can also be used to show a form of optimality for the factorial design. In particular, the levels of experimental inputs may be chosen by the principles of optimal design, where input levels are chosen to optimize a specified criterion. The most common implementation for optimal design involves maximizing the determinant of the Fisher information matrix, commonly referred to as maximizing the *D-optimal* objective function, where "D" represents determinant [1, p. 452], [9], [16, p. 135]. "Fisher Information Matrix" is a summary of facts about the information matrix and its relationship to the linear model here. Optimal design is used in system identification [14, Ch. 6], [22] and can be applied to add rigor and clarify the properties of some of the traditional experimental designs, including factorial designs.

If the noise terms are i.i.d. and normally distributed, then the parameter estimates $\hat{\boldsymbol{\theta}}^{(n)}$ using data from an experimental design that is D-optimal have maximum possible precision in the sense of having the minimum volume confidence regions about $\hat{\boldsymbol{\theta}}^{(n)}$ at a specified probability level. Further, as a consequence of the Kiefer–Wolfowitz equivalence theorem [4, Secs. 17.1–17.2], [9], $\hat{\boldsymbol{\theta}}^{(n)}$ from a D-optimal design provides optimal predictions in the mini-max sense of minimizing the largest possible variance of predictions for future output values z over the domain of possible inputs $\boldsymbol{x}$ when setting $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^{(n)}$ in the linear model. Not all factorial designs are D-optimal. For example, [1, pp. 453–455] shows how a *central-composite design*, a type of factorial design that combines 2^m factorial observations with a star design and center points [1, Sec. 11-4.2], leads to parameter estimates that are less precise than those that follow from a D-optimal design in a problem where the input domain is not a rectangle. In fact, the central-composite design needs more than twice the number of observations of the D-optimal design to achieve parameter estimates with the same accuracy. Discussions on the connection of factorial designs to D-optimality are given in [1, Sec. 11.4.4], [4, Sec. 17.2], [23, Sec. 7.6], and [24, pp. 169–172].

Aside from the geometric and efficiency considerations above, factorial designs may also be D-optimal. Let us demonstrate the optimality of a 2^m factorial design when the noise v_k is i.i.d. The optimality applies for any linear model with main effects and all or some interaction effects. For example, when $m = 3$, the model includes the noise and all or some of the terms associated with β_0 to β_7 shown in (1), so that $1 \leq p \leq 8$ for $m = 3$. We assume that the input domain is a hypercube in $\mathbb{R}^m$ with input factors satisfying $-1 \leq x_{ki} \leq 1$ for all k, i.

With i.i.d. noise, the D-optimal design for linear models of the form (6) is the design that maximizes $\det[\boldsymbol{H}_n^T \boldsymbol{H}_n]$, which is equivalent to maximizing the determinant of the Fisher information matrix for $\hat{\boldsymbol{\theta}}^{(n)}$ or minimizing the determinant of the covariance matrix (7). Because it is assumed that $\boldsymbol{H}_n^T \boldsymbol{H}_n$ is nonsingular, it follows that $\boldsymbol{H}_n^T \boldsymbol{H}_n$ is positive definite, implying from [25, Th. 5, p. 129] that $\det[\boldsymbol{H}_n^T \boldsymbol{H}_n]$ is bounded above by the product of the diagonal entries in $\boldsymbol{H}_n^T \boldsymbol{H}_n$. Consider $n = N2^m$ runs, which corresponds to N replicates of a 2^m factorial design. Because each of the input factors is normalized to be in $[-1, 1]$, each diagonal entry of $\boldsymbol{H}_n^T \boldsymbol{H}_n$ is bounded above by $N2^m$. Hence, $\det[\boldsymbol{H}_n^T \boldsymbol{H}_n] \leq N2^m \times N2^m \times \cdots \times N2^m = (N2^m)^p$. However, for the factorial design, $\boldsymbol{H}_n^T \boldsymbol{H}_n = N2^m \boldsymbol{I}_p$ because the input levels are all chosen at the endpoints ± 1, leading to $\det[\boldsymbol{H}_n^T \boldsymbol{H}_n] = (N2^m)^p$. That is, the factorial design achieves the upper

Fisher Information Matrix

The Fisher information matrix can be defined for many practical identification and estimation problems, including those for which $\boldsymbol{\theta}$ enters nonlinearly or those for which the data are not normally distributed. We provide here a brief discussion of the Fisher information matrix for estimation problems, including the linear regression problem as a special case. Suppose that the probability density, probability mass, or hybrid density/mass function for data $\boldsymbol{Z}_n$ is known, where n is the sample size. Let the probability density, mass, or hybrid function for $\boldsymbol{Z}_n$ be $p_{\boldsymbol{Z}}(\boldsymbol{\zeta}|\boldsymbol{\theta})$, where $\boldsymbol{\zeta}$ is a dummy vector representing the possible outcomes for $\boldsymbol{Z}_n$ and the index n on $\boldsymbol{Z}_n$ is being suppressed for notational convenience. The corresponding likelihood function, denoted $\ell(\boldsymbol{\theta}|\boldsymbol{\zeta})$, satisfies

$$\ell(\boldsymbol{\theta}|\boldsymbol{\zeta}) = p_{\boldsymbol{Z}}(\boldsymbol{\zeta}|\boldsymbol{\theta}).$$

The $p \times p$ Fisher information matrix $\boldsymbol{F}_n(\boldsymbol{\theta})$ for a twice-differentiable log-likelihood function is given by

$$\boldsymbol{F}_n(\boldsymbol{\theta}) \equiv E\left(\frac{\partial \log \ell}{\partial \boldsymbol{\theta}} \frac{\partial \log \ell}{\partial \boldsymbol{\theta}^T}\right) = -E\left(\frac{\partial^2 \log \ell}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\right), \tag{S1}$$

where the expectations are with respect to the data $\boldsymbol{Z}_n$ and the second equality holds when the expectation and derivative in $E(\partial \log \ell / \partial \boldsymbol{\theta})$ may be interchanged [4, Sec. 13.3].

Two key properties of $\boldsymbol{F}_n(\boldsymbol{\theta})$ that are relevant in system identification and other areas of stochastic analysis are asymptotic normality and the Cramér-Rao lower bound. Let $\hat{\boldsymbol{\theta}}_n$ represent an estimate of model parameters $\boldsymbol{\theta}$ from n independent measurements. For example, $\hat{\boldsymbol{\theta}}_n$ might represent a maximum likelihood or Bayesian estimate of parameters in a model that is linear or nonlinear in $\boldsymbol{\theta}$. The asymptotic normality property states that the quantity $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{\text{true}})$ converges in distribution to a Gaussian distribution $N(\boldsymbol{0}, \overline{\boldsymbol{F}}^{-1})$ as $n \to \infty$, where $\overline{\boldsymbol{F}}$ is the limiting average of $\boldsymbol{F}_n(\boldsymbol{\theta})$ over n measurements at $\boldsymbol{\theta} = \boldsymbol{\theta}_{\text{true}}$, that is, $\overline{\boldsymbol{F}} = \lim_{n\to\infty} \boldsymbol{F}_n(\boldsymbol{\theta}_{\text{true}})/n$. A special case is the batch regression estimator associated with (6) and (7), where $\hat{\boldsymbol{\theta}}_n = \hat{\boldsymbol{\theta}}^{(n)}$. The second key property of $\boldsymbol{F}_n(\boldsymbol{\theta})$, the Cramér-Rao bound, states that if $\hat{\boldsymbol{\theta}}_n$ is an unbiased estimator, then $\text{cov}(\hat{\boldsymbol{\theta}}_n) - \boldsymbol{F}_n(\boldsymbol{\theta}_{\text{true}})^{-1}$ is positive semidefinite for all n. In contrast to the asymptotic normality property, the

bound to $\det[\boldsymbol{H}_n^T \boldsymbol{H}_n]$, indicating that the design is a D-optimal solution.

The standard order for representation of the factorial design in Table 1 and Figure S2 is not the unique D-optimal solution because alternative arrangements of $x_{ki} = \pm 1$ can be constructed to yield the same matrix $\boldsymbol{H}_n^T \boldsymbol{H}_n$. For example, $\boldsymbol{H}_n^T \boldsymbol{H}_n$ is unchanged if every component of the leading column has the value –1 instead of 1. It is apparent, however, that if at least one value of x_{ki} is in the interval (–1, 1), then at least one diagonal entry in $\boldsymbol{H}_n^T \boldsymbol{H}_n$ must be strictly below $N2^m$, leading to an upper bound on $\det[\boldsymbol{H}_n^T \boldsymbol{H}_n]$ that is strictly below the value $(N2^m)^p$ achieved with the 2^m factorial design. Hence, every D-optimal solution for linear models of the form (6) with $n = N2^m$ runs must be composed of inputs taking only normalized values ± 1. D-optimal solutions in other settings may also not be unique; see [4, pp. 476–477] for discussion and a linear model example when n is not a multiple of 2^m.

The D-optimality of factorial design rests on the true data-generating mechanism being the same as the assumed linear model in (6). More generally, optimal design for linear or nonlinear models is based on the assumed model being identical to the true system. If the underlying model structure is different from the true system, then the resulting "optimal" design may, in fact, be far from optimal. For example, the choice of inputs under an optimal design may be suboptimal if the assumed model is linear in $\boldsymbol{x}$ and the true system is nonlinear in $\boldsymbol{x}$, a common situation in system identification [15, Ch. 16]. The use of optimal design is, therefore, sometimes advocated for follow-up studies that occur after prior information is used to determine the model form or experiments are conducted to determine whether particular main or interaction effects are statistically significant. This two-step approach of preliminary model analysis followed by optimal design is well recognized in both the statistics literature and control and system identification literature:

> …the performance of optimal designs often depends critically on the validity of the model that drives the search. It [optimal design] can be fruitfully applied in the planning of follow-up experiments because the model is based on objective information obtained from the original experiment. Its use in determining the design for the original experiment is not recommended as a routine practice… [24, p. 172]
>
> It should be noted that the optimal input design will depend on the (unknown) system, so this optimality approach is worthwhile only when good prior knowledge is available about the system. [15, p. 417]

The preliminary model analysis and determination of model form is often carried out by screening experiments with factorial design [16, Sec. 3.2], [24, Sec. 4.4.2]. The optimal design, which is typically not a factorial design, can be implemented in open or closed loop, depending on the nature of the system and the needs of the analysis [26]–[28].

RESPONSE SURFACE METHODOLOGY

Response surface methodology (RSM) is a strategy for building an understanding of the relationship between

Cramér-Rao bound applies with finite sample sizes. Both of these properties show the close connection of $\boldsymbol{F}_n(\boldsymbol{\theta})$ to characterizations of uncertainty in estimates through $\text{cov}(\hat{\boldsymbol{\theta}}_n)$, a connection that is largely responsible for the use of Fisher information in the practice of identification and estimation.

Let us now discuss the form of $\boldsymbol{F}_n(\boldsymbol{\theta})$ in the special case of the linear model of interest (6). Suppose that the n noise terms v_k are both i.i.d. and normally distributed. Then, as shown below, the covariance matrix $\text{cov}(\hat{\boldsymbol{\theta}}^{(n)})$ is equal to the inverse Fisher information matrix for all $n \geq p$; further, the information matrix is not a function of $\boldsymbol{\theta}$. First, note that the n measurements in $\boldsymbol{Z}_n$ are mutually independent. Because there is a common noise variance σ^2 for each measurement, we know that the log-likelihood function is

$$\log \ell(\boldsymbol{\theta} \,|\, \boldsymbol{Z}_n) = c - \frac{1}{2\sigma^2}(\boldsymbol{Z}_n - \boldsymbol{H}_n\boldsymbol{\theta})^T(\boldsymbol{Z}_n - \boldsymbol{H}_n\boldsymbol{\theta})$$
$$= c - \frac{1}{2\sigma^2}(\boldsymbol{Z}_n^T\boldsymbol{Z}_n - 2\boldsymbol{Z}_n^T\boldsymbol{H}_n\boldsymbol{\theta} + \boldsymbol{\theta}^T\boldsymbol{H}_n^T\boldsymbol{H}_n\boldsymbol{\theta}),$$

where the constant term c is related to the variance of the noises, but is not a function of $\boldsymbol{\theta}$. Motivated by the Hessian-based form for the information matrix (S1), we find,

$$\frac{\partial \log \ell(\boldsymbol{\theta} \,|\, \boldsymbol{Z}_n)}{\partial \boldsymbol{\theta}} = -\frac{1}{2\sigma^2}(-2\boldsymbol{H}_n^T\boldsymbol{Z}_n + 2\boldsymbol{H}_n^T\boldsymbol{H}_n\boldsymbol{\theta})$$
$$= \frac{1}{\sigma^2}(\boldsymbol{H}_n^T\boldsymbol{Z}_n - \boldsymbol{H}_n^T\boldsymbol{H}_n\boldsymbol{\theta})$$

and

$$\frac{\partial^2 \log \ell(\boldsymbol{\theta} \,|\, \boldsymbol{Z}_n)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^T} = -\frac{1}{\sigma^2}\boldsymbol{H}_n^T\boldsymbol{H}_n.$$

Therefore, $\boldsymbol{F}_n(\boldsymbol{\theta})^{-1} = \{-E[\partial^2 \log \ell(\boldsymbol{\theta} \,|\, \boldsymbol{Z}_n)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T]\}^{-1} = (\boldsymbol{H}_n^T\boldsymbol{H}_n)^{-1}\sigma^2$, the same as $\text{cov}(\hat{\boldsymbol{\theta}}^{(n)})$ in (7) when considering i.i.d. noise. The Hessian-based form for the information matrix is especially simple in the linear model case because the Hessian is not a function of $\boldsymbol{Z}_n$ or $\boldsymbol{\theta}$. Hence, with linear models, the levels of the input $\boldsymbol{x}$ values embedded in $\boldsymbol{H}_n$ that maximize $\det[\boldsymbol{H}_n^T\boldsymbol{H}_n]$ are the D-optimal solutions. Maximization of the determinant of the general form in (S1) is used for nonlinear problems [4, Sec. 17.4].

input and output variables in a complex process based on a recursive set of localized experiments. RSM began as a tool in the chemical process industry [29] but has evolved to be a tool for the sequential analysis and optimization of complex processes in many different areas. Factorial designs typically play a large role in RSM. This section provides a synopsis of RSM, including discussion on its relationship to various aspects of experimental design. A comprehensive treatment of RSM in the context of system optimization is given in [30], while an example of RSM and factorial design in control systems for automotive engines is given in [12].

Let us focus on the use of RSM for finding an optimal input value $\boldsymbol{x}$ with the aim of minimizing the mean of the response z; hence, the expected value $E(z)$ is the loss function to be minimized. The basic idea is to first collect data at values of $\boldsymbol{x}$ near the currently estimated optimal value $\boldsymbol{x}$. From these runs, a response surface is constructed that describes the local behavior of the process under study. This response surface is typically, but not necessarily, a first- or second-order regression polynomial in the vector of factors $\boldsymbol{x}$. The response surface is used to provide a search direction for moving toward a better value of $\boldsymbol{x}$, and the process repeats itself. The methodology is summarized in "Outline of Response Surface Methodology."

A critical part of RSM is the choice of the experimental points in constructing each response surface, as given in Steps 1 and 2 in "Outline of Response Surface Methodology." The 2^m factorial design, or alternative form of factorial design, is a common way in which the choice of inputs is made [1, Sec. 11.4]. It is often the case that the first few steps of RSM use the main-effects model $z_k = \beta_0 + \boldsymbol{\beta}^T\boldsymbol{x}_k + v_k$, while later steps, requiring greater accuracy, employ the quadratic-input model $z_k = \beta_0 + \boldsymbol{\beta}^T\boldsymbol{x}_k + \boldsymbol{x}_k^T\boldsymbol{B}\boldsymbol{x}_k + v_k$, where β_0, $\boldsymbol{\beta}$, and $\boldsymbol{B}$ represent a scalar, vector, and symmetric matrix to be estimated [1, Sec. 11.1], [24, pp. 390–391]. Hence, in the quadratic case, $\boldsymbol{\theta}$ represents the unique parameters in β_0, $\boldsymbol{\beta}$, and $\boldsymbol{B}$ that are to be estimated in step 2 of RSM. Therefore, to enhance the estimation of the quadratic effects, the later steps require a more complicated choice of input levels than that of the 2^m factorial design.

Figure 5 depicts the RSM process when $m = 2$, illustrating how a simple design may be used in the initial steps of the search process and a refined design in the later steps of the search. The mean output is reduced in the later steps, as reflected in lower values for the neighboring contours of constant response, that is, lower values of constant mean output $E(z)$. D-optimal designs might fruitfully be used in the later steps of the search process as a result of the prior knowledge obtained in the initial steps of the search.

Step 3 in the "Outline of Response Surface Methodology" has $\boldsymbol{x}$ proceed along the path of steepest descent using the most recent response surface as the means of calculating the path. In particular, the search proceeds by changing $\boldsymbol{x}$ by an amount proportional to $-\partial E(z)/\partial \boldsymbol{x}$. For example, with the main-effects model $z = \beta_0 + \boldsymbol{\beta}^T\boldsymbol{x} + v$, the search moves in a direction proportional to $-\boldsymbol{\beta}$. Obvious modifications to this value for $-\partial E(z)/\partial \boldsymbol{x}$ apply with alternative models, such as the quadratic-input model discussed above.

Outline of Response Surface Methodology

Step 0 Initialization: Make a starting guess at the optimal value of $\boldsymbol{x}$.

Step 1 Collect responses z from several values of $\boldsymbol{x}$ in the neighborhood of the current value of $\boldsymbol{x}$. To enhance the fitting process, use experimental design to determine the values of $\boldsymbol{x}$ with which to collect measurements z.

Step 2 From the pairs $\boldsymbol{x}$, z in Step 1, fit a regression—that is, response surface—model over the region around the current value of $\boldsymbol{x}$.

Step 3 From the response surface in Step 2, estimate a path of steepest descent in the domain for $\boldsymbol{x}$ within $\mathbb{R}^m$.

Step 4 Perform a series of runs at values of $\boldsymbol{x}$ along the path of steepest descent until no additional improvement in the process is obtained. As a consequence of noise in the measurements z, statistical tests are typically required to determine when the process $E(z)$ is no longer improving [30, Ch. 6]. The terminal value of $\boldsymbol{x}$ along the path represents the new estimate of the best vector of factor levels.

Step 5 Go to Step 1 and repeat the process until the optimal value of $\boldsymbol{x}$ is obtained. Second-order (Hessian matrix) analysis can be used to help verify that the solution is at least a local minimum of $E(z)$.

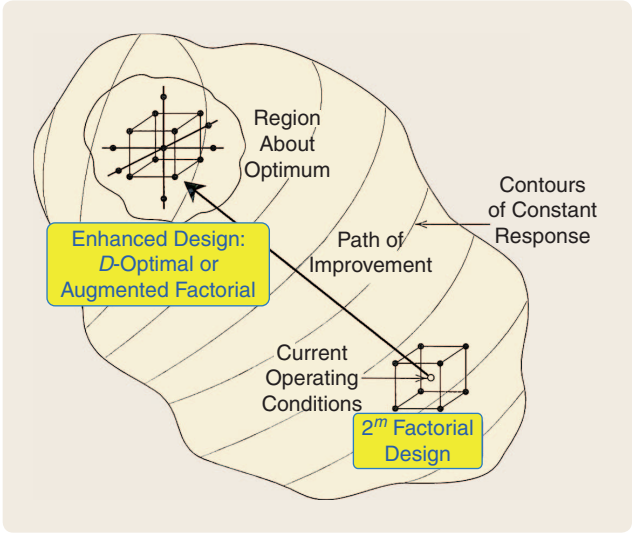

FIGURE 5 Conceptual illustration of the response surface methodology when $m = 2$. The plot depicts a 2^m factorial design when $\boldsymbol{x}$ is far from the optimum and an enhanced design when $\boldsymbol{x}$ is near the optimum, reflecting a greater system understanding. Candidate designs near the optimum include a D-optimal design or an augmented factorial design such as a 2^m factorial design with strategically chosen additional points. The enhanced design near the optimum may be both a D-optimal design and a type of factorial design. (Figure 5 is a modified version of Figure 11.3 in [1]. Reproduced with permission of John Wiley & Sons, Inc.)

Responses are observed along the search path until there is no significant improvement in the response (step 4). Of course, because we are working with only an estimate of $\boldsymbol{\beta}$, not an underlying "true value," the gradient used in the steepest descent is only an estimate of the true gradient $\partial E(z)/\partial \boldsymbol{x}$. Therefore, the search process is a version of the *stochastic gradient* algorithm, which is a special case of stochastic approximation (SA) [4, Ch. 5].

Although RSM has a long history of success, there is no guarantee of greater efficiency, in the sense of needing fewer experimental runs to reach a solution, than with alternative stochastic optimization methods [4, p. 488], [30, p. 10 and Ch. 13]. In particular, SA in the form of stochastic gradient methods or gradient-free methods [4, Chaps. 5–7] can also be useful for minimizing $E(z)$ without the process of localized curve fitting inherent in RSM. As with RSM, SA methods are based on noisy observations of the criterion $E(z)$, with values of z as noisy measurements of $E(z)$, or noisy measurements of the gradient $\partial E(z)/\partial \boldsymbol{x}$.

STATISTICAL ANALYSIS, EXTENSIONS, AND CONNECTIONS TO STOCHASTIC OPTIMIZATION

The discussion above largely suppresses the issue of statistical significance for the results of a factorial experiment. That is, we wish to determine whether various observed nonzero main and interaction effects are statistically significant. This issue is often addressed using standard methods from ANOVA and linear regression to carry out the multiple comparisons in a meaningful way when it can be assumed that the noise terms v_k are i.i.d. and normally distributed [1, Sec. 5.3 and Sec. 5.4], [2, Ch. 4], [3, Ch. 13]. The i.i.d. assumption allows for an estimate of the noise variance σ^2 to be calculated based on all of the data over the range of input values. Tests on main and interaction effects can be performed from a single replicate of the factorial experiment or, better, from multiple randomized-input replicates. Standard statistical F-tests or t-tests form the basis for hypothesis tests and calculation of probability values, commonly called P-values. It is important, however, to verify that the stringent assumptions of the ANOVA method hold if ANOVA is used to draw conclusions about main and interaction effects. Some of the tests available for checking assumptions are summarized in [3, pp. 317–321]. See "Factorial Design for a Circuit Board, Part 2: Estimation and Statistical Significance" for an example of the application of standard statistical methods in 2^m factorial design.

Although the ANOVA and standard regression-based testing methods mentioned above are the preferred approach for hypothesis testing when the noise distribution has the required i.i.d. structure, alternative testing methods apply when the noise distribution is more complex. One common circumstance for the noise to not be i.i.d is when the noise depends on the input $\boldsymbol{x}$. Examples of such dependence in problems related to discrete-event systems are given in [8, pp. 627–635]. If the distributions of the noise

Factorial Design for a Circuit Board, Part 2: Estimation and Statistical Significance

The main and two-way interaction effects for the three factors of interest in the resistor problem discussed in "Factorial Design for a Circuit Board, Part 1: Formulation and Data" are computed from the experimental results in Table S1. Three-way interaction is not considered because of a lack of engineering interest and the limited quantity of data. Hence, $p = 7$. The noise is assumed to be i.i.d., indicating that inference can be carried out using standard analysis of variance or linear regression. While it is recognized that the i.i.d. assumption may be suspect, the overall noise level is small, making the assumption less critical here than in applications with larger levels of noise. Table S2 shows the results of a regression analysis using an input matrix of the form (8) with the last column removed because three-way interactions are not considered. Thus, $\boldsymbol{H}_n^T\boldsymbol{H}_n = \boldsymbol{H}_8^T\boldsymbol{H}_8 = 8\boldsymbol{I}_7$ in the regression calculations. The estimate of σ, the standard deviation of the noise, is 1.47, consistent with the noise level being small given that the observed values of z are from -0.01 to 120.6.

Table S2 shows that the two-sided P-values for the parameters associated with inputs s, sr, and sh are all at least 0.5 while the P-values for the additive constant and parameters associated with inputs r, h, and rh are all only slightly more than 0.01 (the P-value is the probability of a future estimate of the indicated β_i having a magnitude greater than the observed estimate in Table S2 under an assumption that the true value of the parameter is zero). Hence, the results in Table S2 provide no evidence that the higher soldering temperature is a contributor to the increased resistance, but the analysis suggests that the resistor batch coupled with humidity level appears to be a contributor to the increase. The individual P-values for the β_i associated with r, h, and rh are sufficiently small to suggest nonzero true β_i despite the multiple comparisons associated with $p > 1$.

This example illustrates the power of designed experiments to determine whether interactions are statistically significant, in this case the interaction of resistor batch and humidity. Interactions are often missed with one factor at a time experiments. Because the experiment revealed that one of the likely causes of increased resistance is high humidity, which is irrelevant outside of the atmosphere, and because the contribution of the choice of resistor batch to changes in resistance was not expected to cause significant problems once the spacecraft was in orbit, it was decided to proceed with the launch of MESSENGER using the circuit boards already on the spacecraft. As noted in Part 1 of this example, the launch occurred successfully in August 2004. Furthermore, the circuit boards have functioned properly throughout the mission.

TABLE S2 Regression estimates and tests of statistical significance for resistor data. Estimates and *P*-values in bold font indicate statistical significance at testing levels of slightly more than 1%. Proper simultaneous interpretation of the *P*-values for more than one parameter must include the multiplicity effect.

Parameter	Input Factor	Parameter Estimate	*P*-value
β_0	Constant	**30.38**	**0.0109**
β_1	s	−0.058	0.929
β_2	r	**−29.34**	**0.0113**
β_3	h	**30.39**	**0.0109**
β_4	sr	0.52	0.5
β_5	sh	−0.053	0.935
β_6	rh	**−29.34**	**0.0113**

terms depend on $\boldsymbol{x}$, or are otherwise non-i.i.d., and the covariance matrix $\mathrm{cov}(\boldsymbol{V}_n)$ is known, at least to a scalar multiple, then a generalized least-squares method can be used to produce an optimal least-squares estimate [2, pp. 125–127], [4, p. 71]. Most of the standard statistical techniques for regression analysis continue to apply in such a generalized setting.

In some cases, $\mathrm{cov}(\boldsymbol{V}_n)$ is unknown to within a scalar multiple, and thus generalized least squares is not appropriate. A way to address the question of statistical significance in such settings is to collect output from N independent replicates of the 2^m factorial design and obtain N independent values for a main or interaction effect being tested [8, pp. 627–636]. It is then straightforward to calculate confidence intervals and carry out standard statistical t-tests for the given effect based on the sample means and variances of the N independent calculations for the given effect together with the t-distribution with $N-1$ degrees of freedom. Further, for N reasonably large, the t-test can be reliably implemented without assuming that the individual effect calculations going into the sample mean calculation are normally distributed [4, Appendix B]. For example, a test for the significance of main effect i would rely on the sample mean and variance of N calculations of expression (3), together with the t-distribution, without necessarily assuming that the observations z_k are normally distributed. Note, however, that care is needed in making statements about simultaneous statistical significance of multiple effects. That is, as more effects are tested, it is increasingly likely that at least some effects are going to show statistical significance due to inherent random variation. Methods for compensating for such multiple comparisons are discussed in [1, Sec. 3.5], [2, Sec. 7.5], [4, Ch. 12], and [8, Sec. 9.7]. Alternative methods for dealing with complex relationships between the noise variance and $\boldsymbol{x}$ are given in [5, Sec. 3.4]; these methods

involve the collection of additional system data or the running of bootstrap or jackknife simulations.

The 2^m factorial design has many variations and extensions. If m is even moderately large, the 2^m factorial design may require an excessive number of input–output runs. For example, with $m = 12$, which is modest by simulation standards, then one full-factorial experiment requires $2^{12} = 4096$ runs. If, in a simulation setting, $N = 100$ replicates of the experiment are required to obtain meaningful statistical results, and each input–output run takes, say, 6 s (not large by serious simulation standards), then the full study would require nearly one month of execution time. To address this cost issue, fractional factorial designs are often used. In such designs, a subset of size 2^{m-q} of the input combinations is used where $1 \leq q < m$. The main issue, of course, is how to choose the value of q and the specific input combinations. This subject is covered extensively in the classical experimental design literature [1, Ch. 8], [2, Sec. 11.7], and [3, Ch. 5].

The 2^m factorial design is oriented to cases where each component of $\boldsymbol{x}$ enters directly or by means of interactions. If, in fact, there is reason to believe that quadratic or other nonlinear input effects are present, such as when the model includes terms x_{ki}^2 for at least one i, then it is necessary to obtain observations at inputs other than the two $+/-$ values of a 2^m factorial design to gain information about the curvature. A common way is to preserve the basic structure of the 2^m factorial design but augment the design with additional runs at appropriate points. Perhaps the simplest method is to place additional runs at the m-dimensional point $\mathbf{0}$ together with other selected points, where the ith individual component 0 represents an intermediate point between $-$ and $+$ on the chosen scale for the ith component of $\boldsymbol{x}$. In particular, the above-mentioned central-composite design places the additional points in the center and along the coordinate axes corresponding to each factor, as in the star design. The central-composite design is the most widely used method for second-order models in the inputs as a consequence of the efficient use of experimental resources [1, p. 440], [24, p. 412]. An alternative is to use a 3^m factorial design, where runs are carried out at all possible combinations of -1, 0, and 1 for the input factors. The 3^m design, however, is typically less efficient than the central composite design in terms of the number of runs required to detect curvature at a given level of statistical significance [1, Ch. 9].

Experimental design is not the only area where the idea of simultaneous changes in input variables is useful. For example, in stochastic optimization, the simultaneous perturbation SA (SPSA) algorithm is also based on changing all terms at once instead of the one-at-a-time changes of the classical finite difference SA method [4, Ch. 7], [31]. A fundamental difference in SPSA and factorial design, however, is that the simultaneous changes in SPSA are random in contrast to the deterministic changes in factorial design. A further difference is that SPSA is based on a recursive process involving a small number of output measurements, independent of the problem dimension, at each step, while factorial design involves a collective acquisition of the measurements as part of one experiment, with the number of measurements growing with the problem dimension. Despite these differences, large efficiency gains are possible in both SPSA and factorial design in terms of the total number of measurements needed to achieve a given accuracy in the final model estimate relative to the number of measurements needed in one-at-a-time changes to achieve the same accuracy. In particular, the efficiency gain in both SPSA and, as shown in Figure 3, factorial design is proportional to the problem dimension.

CONCLUSIONS

"A well-designed experiment is an efficient method of learning about the world" [16, p. vii].

Factorial design is one of the most common and powerful strategies for choosing the values of inputs in experiments. Relative to alternative strategies, such as classical one-at-a-time changes, factorial design with its simultaneous changes in multiple factors typically provides greater insight into a system with a lower number of required experimental runs. Using the optimal design approach popular in system identification [9], factorial designs can sometimes be shown to provide the best possible choice of input levels for purposes of extracting information from a set of tests.

It is also necessary, of course, to recognize certain limitations. Most standard experimental design approaches, including factorial design, are based on the assumption that the underlying model is linear in the parameters $\boldsymbol{\theta}$ being estimated. Such an assumption is often not true. In fact, a simple illustration of factorial design working more poorly than a straightforward engineering approach is given in [32] in the context of a paper helicopter. This system has components within $\boldsymbol{\theta}$ that enter quadratically due to aerodynamic effects. A helicopter constructed from the optimization of a model built from engineering first principles significantly outperforms a different helicopter built from optimizing a model constructed using factorial design. Therefore, when there is a known significant nonlinearity in the system that is best modeled by means of nonlinear contributions of the model parameters, it may be better to use an engineering first-principles analysis or an optimal design method appropriate for nonlinear processes [4, Sec. 17.4], [16, Chaps. 17, 18]. Hence, as with all practical algorithms and methods, it is essential to use factorial design with an appropriate understanding of the system being studied.

Factorial design is an effective strategy for helping to determine the effect of input factors on system output

when physical knowledge is nonexistent or limited. Factorial methods are often used in screening experiments in the early stages of design and experimentation to determine the relative importance of particular inputs or specific interactions between inputs. Factorial design may also be used in the later stages for enhanced identification and parameter estimation. The large number of successful applications of factorial design in industry, government, and elsewhere is a testament to the power of this strategy.

ACKNOWLEDGMENTS

This work was partially supported by U.S. Navy Contract N00024-03-D-6606. Prof. Dariusz Uciński of the Institute of Control and Computation Engineering, University of Zielona, Poland, provided several useful references, and Dr. Manuel Uy of the Johns Hopkins University, Applied Physics Laboratory, provided valuable comments from a practitioner's perspective. Dr. Uy also provided the essential information for the example pertaining to increased resistance in circuit boards on the MESSENGER spacecraft. The author also thanks an anonymous referee for many insightful comments.

AUTHOR INFORMATION

James C. Spall (james.spall@jhuapl.edu) is a member of the principal professional staff at the Johns Hopkins University (JHU), Applied Physics Laboratory, Laurel, Maryland, and is a research professor in the JHU Department of Applied Mathematics and Statistics, Baltimore, Maryland. He is also chair of the Applied and Computational Mathematics Program within the JHU Engineering and Applied Science Programs for Professionals. He has published in the areas of statistics and control and holds two U.S. patents (both licensed) for inventions in control systems. He is the editor and coauthor of *Bayesian Analysis of Time Series and Dynamic Models* [Marcel Dekker (now CRC Press), 1988] and the author of *Introduction to Stochastic Search and Optimization* (Wiley, 2003). He was an associate editor and associate editor at large for *IEEE Transactions on Automatic Control*, where he is currently a senior editor. He is also a contributing editor for *Current Index to Statistics*. He was the program chair for the 2007 IEEE Conference on Decision and Control and is a Fellow of IEEE. He can be contacted at John Hopkins University, Applied Physics Laboratory, 11100 Johns Hopkins Road, Laurel, MD 20723 USA.

REFERENCES

[1] D. C. Montgomery, *Design and Analysis of Experiments*, 7th ed. Hoboken, NJ: Wiley, 2009.

[2] K. Hinkelmann and O. Kempthorne, *Design and Analysis of Experiments, Vol. 1: Introduction to Experimental Design*, 2nd ed. Hoboken, NJ: Wiley, 2008.

[3] T. B. Barker, *Quality by Experimental Design*, 3rd ed. Boca Raton, FL: CRC, 2005.

[4] J. C. Spall, *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*. Hoboken, NJ: Wiley, 2003.

[5] J. P. C. Kleijnen, *Design and Analysis of Simulation Experiments*. New York: Springer-Verlag, 2008.

[6] K. M. Passino, *Biomimicry for Optimization, Control, and Automation*. London: Springer-Verlag, 2004.

[7] J. P. C. Kleijnen, S. M. Sanchez, T. W. Lucas, and T. M. Cioppa, "State-of-the-art review: A user's guide to the brave new world of designing simulation experiments," *INFORMS J. Comput.*, vol. 17, no. 3, pp. 263–289, 2005.

[8] A. M. Law, *Simulation Modeling and Analysis*, 4th ed. New York: McGraw-Hill, 2007.

[9] L. Pronzato, "Optimal experimental design and some related control problems," *Automatica*, vol. 44, pp. 303–325, 2008.

[10] K. Lee and J. Kim, "Controller gain tuning of a simultaneous multi-axis PID control system using the Taguchi method," *Control Eng. Pract.*, vol. 8, pp. 949–958, 2000.

[11] D. D. Kukolj, M. T. Berko-Pusic, and B. Atlagic, "Experimental design of supervisory control functions based on multiplayer perceptrons," *Artif. Intell. Eng. Des., Anal. Manuf.*, vol. 15, pp. 425–431, 2001.

[12] D. Popović, M. Janković, S. Magner, and A. R. Teel, "Extremum seeking methods for optimization of variable cam timing engine operation," *IEEE Trans. Control Syst. Technol.*, vol. 14, no. 3, pp. 398–407, 2006.

[13] J. C. Spall and J. P. Garner, "Parameter identification for state-space models with nuisance parameters," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 26, no. 6, pp. 992–998, 1990.

[14] G. C. Goodwin and R. L. Payne, *Dynamic System Identification: Experiment Design and Data Analysis*. New York: Academic, 1977.

[15] L. Ljung, *System Identification—Theory for the User*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall PTR, 1999.

[16] A. C. Atkinson, A. N. Donev, and R. B. Tobias, *Optimum Experimental Designs, with SAS*. London: Oxford Univ. Press, 2007.

[17] D. B. Duncan and S. D. Horn, "Linear dynamic recursive estimation from the viewpoint of regression analysis," *J. Amer. Statist. Assoc.*, vol. 67, pp. 815–821, 1972.

[18] J. C. Spall and K. D. Wall, "Asymptotic distribution theory for the Kalman filter state estimator," *Commun. Statist.: Theory and Methods*, vol. 13, no. 16, pp. 1981–2003, 1984.

[19] F. A. Aliev and L. Ozbek, "Evaluation of convergence rate in the central limit theorem for the Kalman filter," *IEEE Trans. Automat. Contr.*, vol. 44, no. 10, pp. 1905–1909, 1999.

[20] V. Geramita, J. M. Geramita, and J. S. Wallis, "Orthogonal designs," *Linear Multilinear Algebra*, vol. 3, no. 4, pp. 281–306, 1976.

[21] L. Raktoe, A. Hedayat, and W. T. Federer, *Factorial Designs*. New York: Wiley, 1981.

[22] E. Walter and L. Pronzato, "Qualitative and quantitative experiment design for phenomenological models—A survey," *Automatica*, vol. 26, pp. 195–213, 1990.

[23] A. C. Atkinson and A. N. Donev, *Optimum Experimental Designs*. London: Oxford Univ. Press, 1992.

[24] C. F. J. Wu and M. Hamada, *Experiments: Planning, Analysis, and Parameter Design Optimization*. New York: Wiley, 2000.

[25] R. Bellman, *Introduction to Matrix Analysis*, 2nd ed. New York: McGraw-Hill, 1970.

[26] H. Hjalmarsson, "From experiment design to closed-loop control," *Automatica*, vol. 41, no. 3, pp. 393–438, 2005.

[27] D. Uciński, *Optimal Measurement Methods for Distributed Parameter System Identification*. Boca Raton, FL: CRC, 2005.

[28] J. C. Agüero and G. C. Goodwin, "Choosing between open and closed loop experiments in linear system identification," *IEEE Trans. Automat. Contr.*, vol. 52, no. 8, pp. 1475–1480, 2007.

[29] G. E. P. Box and K. B. Wilson, "On the experimental attainment of optimum conditions," *J. R. Statist. Soc., Series B*, vol. 13, pp. 1–38, 1951.

[30] E. Del Castillo, *Process Optimization: A Statistical Approach*. New York: Springer-Verlag, 2007.

[31] J. C. Spall, "Multivariate stochastic approximation using a simultaneous perturbation gradient approximation," *IEEE Trans. Automat. Contr.*, vol. 37, no. 3, pp. 332–341, 1992.

[32] H. Annis, "Rethinking the paper helicopter: Combining statistical knowledge and engineering knowledge," *Amer. Statist.*, vol. 59, pp. 320–326, 2005.