Non-orthogonality in test design: practical relevance of the theoretical concept in terms of regression quality and test plan efficiency

Philipp Mell[a], Marco Arndt[a], and Martin Dazer[a]
[a]Institute of Machine Components, University of Stuttgart, Stuttgart, Germany,
philipp.mell@ima.uni-stuttgart.de

Abstract: Whenever scientific problems aren't well understood in their physical properties or cannot be solved analytically, the approach of design of experiments (DoE) is a powerful alternative. Yet, many DoE approaches are mathematically derived and underly assumptions and restrictions which might be hard or even impossible to be met in practice. Therefore, numerous research gaps regarding the practical implementation of DoE test plans remain.
One typical requirement is that the experimental design has to be orthogonal. This condition demands that the investigated factors be set exactly to the given factor levels, which is usually impossible. The question arises, how crucial the impact of a particular deviation from the ideal orthogonal design is. Until now, this has only been assessed qualitatively, but not quantitatively. Different metrics known in literature could be utilized to measure the non-orthogonality of a test plan, which are presented in this paper. The consequences of non-orthogonality are evaluated on the basis of two central quantities of a DoE test plan: First, the power, which shows how likely an existing effect is to be identified. Second, the accuracy of the estimated model parameters resulting from the regression model developed from the test results. The scope of this paper is the assessment of these two quantities for typical deviations from a perfectly orthogonal full factorial test plan, allowing a transfer between theoretical requirements and practical use. In the long term, the results can be utilized to make test plans more efficient by suggesting which cost-reducing types of non-orthogonality still produce satisfying results. In order to achieve this goal, a parameter study for two hypothetical two- and three-dimensional systems is performed in a Monte Carlo simulation. For both the orthogonal and non-orthogonal test plans, linear regression and significance analysis is applied. The changes in test power and regression accuracy allow to assess, how crucial the different types of non-orthogonality are. Also, the results are compared with two deliberately chosen non-orthogonality measures, to see which of them serves best in predicting the practical value of a DoE test plan.

1. INTRODUCTION

Gaining empirical evidence from experimental data is one of the core constituents of scientific research, be it in natural science or in social sciences. By planning, conducting, observing and evaluating experiments on the system of interest, new insights into the system's properties and its dependency on influencing parameters are gained. It is therefore of universal interest to optimize the design of such experiments. Arguably the most prominent pioneer of defining and formalizing basic principles towards optimized experiments is R. A. Fisher, according to whose primary publication this subject area is labeled *Design of Experiments* (DoE) [1]. In this sense, the term DoE refers to the deliberate planning of experiments under consideration of the subsequent evaluation by using statistical methods. The aim of this approach is to either maximize the quality of the results (e.g. the detectability or accuracy of influences) or minimize the effort (e.g. the cost, expenditure of time or global warming potential) for a particular use case. Another possible aim, combined of these two fundamental ideas, is an *efficient* experimental design, striving to optimize the ratio between quality and effort.
The power of DoE is based on a set of fundamental principles developed by Fisher, e.g. randomization, factorial designs or blocking [1-7]. Among these, the factorial designs stand out in particular. They involve the simultaneous, systematic modification of influencing factors on all of their possible values. Before that, the approach of modifying only *one factor at a time* (OFAT) was predominant, leading to

inefficient test plans which ignore possible interactions [3,4]. Factorial designs have been adapted and extended by numerous authors, leading to plentiful different test plans. Some common examples are Plackett-Burmann designs, central composite designs, and Box-Behnken designs [2-5,7].
However, one of the DoE principles is rarely discussed due to its highly mathematical foundation: orthogonality of the conducted test plan. Orthogonality guarantees both the absence of correlation between the parameter estimates as well as a minimum standard error of the regression model coefficients [8,9]. Perfect orthogonality, however, is unachievable in practical experiments. For example, random deviations in the setting of the influencing factors, the systematic offset of an influencing factor in one or more factor combinations, or a completely left out factor combination may be encountered in actual experiments. The reason for this are either timely, monetary or physical constraints. The question arises, how such imperfectness regarding the test plan orthogonality affects the quality of the test results. This question has been tackled in research only qualitatively [3-7] or for simple example cases without consideration of the detectability of effects [2]. A profound quantitative analysis in the face of practical cases of application is therefore necessary.
In the face of ever-increasing time and cost pressure, the opposite question may be asked as well: how high of a deviation from an orthogonal design is admissible, given certain demands on the quality of the test result? Perspectively, the answer to these question leads from general orthogonal to efficient test plans, which are optimal for the individual use case. Therefore, this paper aims to

(i) present an introduction into the concept of test plan (non-)orthogonality;
(ii) give an overview over possible non-orthogonality measures;
(iii) define and simulate realistic scenarios of deviations from perfect orthogonality and assess the consequences on the quality of the test results.

Since this paper merely constitutes a fundamental work, only linear effects are considered. Therefore, two-stage full factorial test plans are used as a basis. Results are generated exclusively simulatively by using the Monte Carlo method. In the following, the theoretic background – including hypothesis testing and regression analysis – is explained in Section 2. Section 3 follows with a focus on (non-) orthogonality, before Section 4 defines a parameter study and presents its results.

2. THEORETIC BACKGROUND

This section serves to give a brief overview over the basic terms used in following sections and the scope of systems covered in this paper. Subsequently, it gives the necessary background on hypothesis testing and DoE evaluation.

2.1 System definition

DoE serves as a method to systematically configure and evaluating experimental studies, ultimately aiming to mathematically describe a system better than before. A linear system's output y with interactions only between two factors – as governed in this paper – can be described by the equation

$$y = c_0 + \sum_{i=1}^{m} c_i x_i + \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} c_{ij} x_i x_j + \varepsilon, \tag{1}$$

where x_i is the i-th of the m input variables, c_i is the coefficient of the i-th parameters influence, c_0 is the intercept (i.e., the system output in the origin of the parameter space and therefore the global mean output), c_{ij} is the coefficient of the interaction between the i-th and the j-th input variable and ε is the error term.
The sole purpose of any experiment and therefore DoE is to find out

a) which of the considered parameters has an influence on the output variable, and, if applicable,
b) how high this influence is by implementing a regression analysis.

The theoretical background for the answer to these two questions is presented in Section 2.2 and Section 2.3, respectively.

2.2 Hypothesis testing

The first of the two goals mentioned in the previous section is accomplished on the basis of hypothesis testing. For each influence and interaction, two complementary hypotheses are formulated:

H_0: the null hypothesis, which states that the parameter does have no influence on the output;

H_1: the alternative hypothesis, stating that is does have an influence on the output.

This idea is applied to the system from eq. (1): the output variable is seen as a stochastic variable due to the unexplainable influence of the random error ε. For this reason, the output y itself is a stochastic variable. For each of the input variables x_i, the null hypothesis is that the mean output $\bar{y}$ does not change when x_i changes from one level to another. The alternative hypothesis is that it does, no matter in which direction. Defining $\bar{y}_{i+}$ as the mean output on the higher level of parameter x_i and $\bar{y}_{i-}$ as the mean output on the lower level (see Figure 1) allows to formalize these hypotheses:

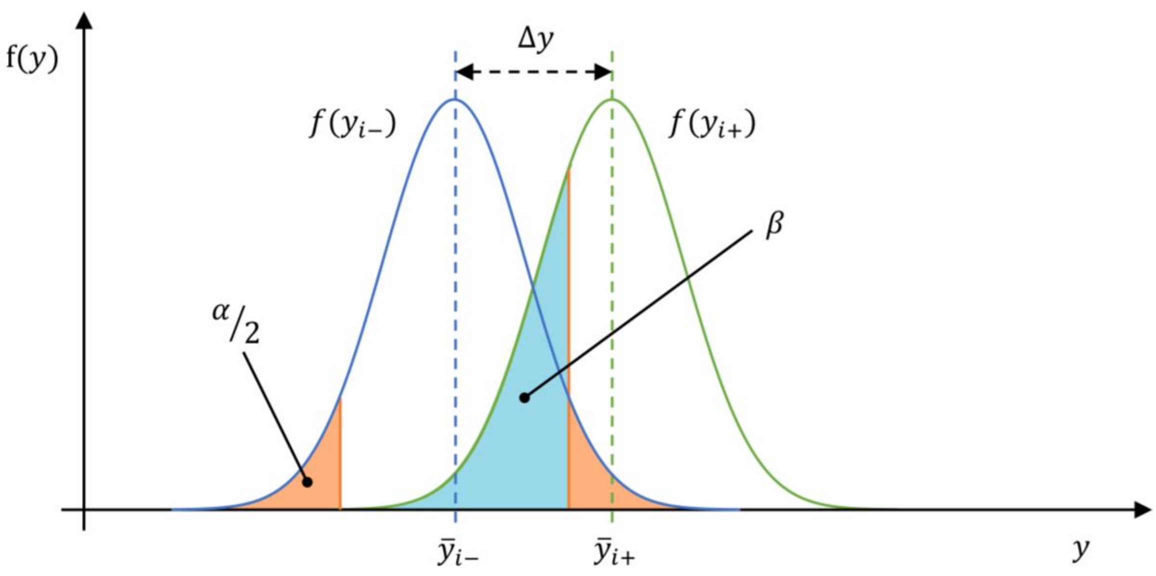
$$H_0: \ \bar{y} \neq f(x_i) \tag{2}$$

$$H_1: \ \bar{y} = f(x_i) \tag{3}$$

Since y is a stochastic variable, a difference in the output, Δy, on two levels of x_i is not necessarily due to an influence of x_i on the mean output $\bar{y}$, but could also be due to chance (i.e., the variance σ^2). Falsely drawing the conclusion that x_i has an influence on y – even though it actually doesn't – is called a *type I error* or *false positive*. The risk of such an error can never be completely eliminated, since any change in the output could be due to chance. A sufficiently high difference Δy though is highly unlikely to be from chance. Therefore, inputs with high influence – leading to higher output differences Δy – and repeated experiments both make it easier to detect an influence through the experimental data.

To formalize the risk of falsely attributing an influence to a parameter x_i, the highest admissible probability of a type I error is defined before the experiment. The according value, α, is called the significance level (see Figure 1).

Figure 1: Probability density functions of an output quantity y with mean $\bar{y}$, when the i-th parameter is on the low ($-$, blue) or high ($+$, green) level. The null hypothesis, H_0, is that the change of the i-th parameter from low to high level doesn't systematically affect the mean output $\bar{y}$. For a demanded significance level α, this is assumed as long as the mean $\bar{y}_{i+}$ is so close to $\bar{y}_{i-}$ that it doesn't touch the orange area. This is not the case here. Hence, the null hypothesis is rejected and the effect of parameter i is significant (alternative hypothesis H_1). The turquoise area shows the type II error probability β; the complementary area under the green distribution is the test power $1 - \beta$.



The probability to observe a certain output difference Δy, although H_0 holds, is called the p-value. An effect is said to be statistically significant, if it's p-value does not exceed the imposed limit α and the null hypothesis H_0 is therefore rejected. This is equivalent to saying that the risk of falsely assuming an influence is equal to p. The other possible error an experiment might yield is falsely denying an effect although it actually exists. This is called a *type II error* or *false negative*. The probability of committing such a mistake is denoted β. Naturally, this value increases when α is decreased, since a lower risk of a false positive result (i.e., a lower α) demands the effect to be more clearly visible. Effects below that limit, even though systematical, will then be rejected as non-significant, yielding a higher share of false negatives (β). Other than α, β cannot be set beforehands, since it depends on both α and the effect size Δy. Rather, it is calculated after the experiment. Its complementary, $(1-\beta)$, is called the test power: the chance of correctly detecting an existing influence. The test power can be used as a quality measure when comparing different test plans.

2.3 Regression analysis

The second of the two goals mentioned in Section 2.1 is to estimate the model coefficients c_i of the system (cf. eq. (1)). This purpose is usually served by applying multivariate linear regression with the ordinary least squares (OLS) method. The reason for this is that under the Gauss-Markov theorem – which is assumed to hold here [8,9] –, OLS delivers the best linear unbiased estimator. The problem solved by multivariate regression is to find an estimate for the coefficients c_i of a problem of the form

$$y = c_0 + \sum_{i=1}^{k} c_i z_i + \varepsilon \overset{z_0=1}{=} \sum_{i=0}^{k} c_i z_i + \varepsilon, \tag{4}$$

where c_i are coefficients for the k input variables z_i and ε is the error term. The coefficient c_0 is the mean output. Therefore, a total of $p = k + 1$ coefficients are to be estimated. As can be seen from the similarity between eqs. (1) and (4), the DoE problem can be transformed by assigning the m system input quantities x_i and the derived quantities (e.g., the interactions $x_i x_j$) to the OLS input variables z_i. A more convenient representation of eq. (4) in matrix form is obtained when defining the constant $z_0 = 1$ as part of the inputs z_i:

$$\boldsymbol{y} = \boldsymbol{Zc} + \boldsymbol{e}. \tag{5}$$

Here, $\boldsymbol{y}$ is a $(n \times 1)$ column vector, containing the observations of the output from n conducted experiments (with $n > p$), $\boldsymbol{c}$ is the $(p \times 1)$ column vector of coefficients and $\boldsymbol{e}$ is the $(n \times 1)$ row vector of residuals. $\boldsymbol{Z}$ is the $(n \times p)$ data matrix, containing $(n \times 1)$ column vectors $\boldsymbol{z_i}$ with the $\boldsymbol{i}$-th parameters value in each experiment; $\boldsymbol{z_0}$ is a vector of 1's. The OLS estimate of the coefficients, $\hat{\boldsymbol{c}}$, is then computed via

$$\hat{\boldsymbol{c}} = \boldsymbol{Z_L^{-1} y}, \tag{6}$$

with $\boldsymbol{Z_L^{-1}}$ being the $(p \times n)$ pseudoinverse (or generalized left inverse) of $\boldsymbol{Z}$. It is defined through

$$\boldsymbol{Z_L^{-1}} = \boldsymbol{M^{-1} Z^T} = \left(\boldsymbol{Z^T Z}\right)^{\boldsymbol{-1}} \boldsymbol{Z^T}, \tag{7}$$

where $[\cdot]^{\boldsymbol{T}}$ denotes a transposed matrix and $[\cdot]^{\boldsymbol{-1}}$ denotes a matrix inverse. The pseudoinverse exists if and only if the number of observed different parameter combinations n is greater than the number of coefficients in the system model k.

With the estimated outputs $\hat{\boldsymbol{y}} = \boldsymbol{Z}\hat{\boldsymbol{c}}$, the residuals $\hat{\boldsymbol{e}}$ can be calculated as

$$\hat{\boldsymbol{e}} = \boldsymbol{y} - \hat{\boldsymbol{y}} = \boldsymbol{y} - \boldsymbol{Z}\hat{\boldsymbol{c}}. \tag{8}$$

It must be noted that the true errors $\boldsymbol{e}$ remain unknown, yet the residuals $\hat{\boldsymbol{e}}$ are utilized as an estimate for them. The unknown error variance σ^2 can be estimated from the residuals with

$$\hat{\sigma}^2 = \frac{\hat{\boldsymbol{e}}^T \hat{\boldsymbol{e}}}{n-p}, \tag{9}$$

where the denominator $(n-p)$ marks the statistical degrees of freedom. The square root of $\hat{\sigma}^2$, $\hat{\sigma}$, is the standard error of the regression. To increase the accuracy of the estimated coefficients, their variances have to be decreased. The variances of the estimated coefficients depend on the system-based

variance σ^2 and the inverse of the moment matrix $\boldsymbol{M}$. Therefore, two ways of increasing the estimation accuracy are possible: On the one hand, one can reduce the overall variance σ^2, which ultimately means to address the sources of the errors ε, e.g. inaccurate measurement or neglected influences. Such considerations lie without the scope of this paper. On the other hand, one can aim to decrease the inverse of the moment matrix $\boldsymbol{M}$, which can only be achieved by manipulations of the data matrix $\boldsymbol{Z}$. This connects back to the previous section, as the data matrix contains the factor combinations which are set in the test plan. Therefore, the fundamentals of the OLS presented here give a mathematic reason for using DoE methods to optimize test plan.
For further evaluation of the conducted experiments, the so-called *t statistic* of each estimated coefficient is calculated. This is done by dividing each parameters OLS estimate through its standard deviation [9]. From the *t statistic*, the p-value as defined in Section 2.2 can be calculated by applying the cumulative distribution function of Student's t-distribution.

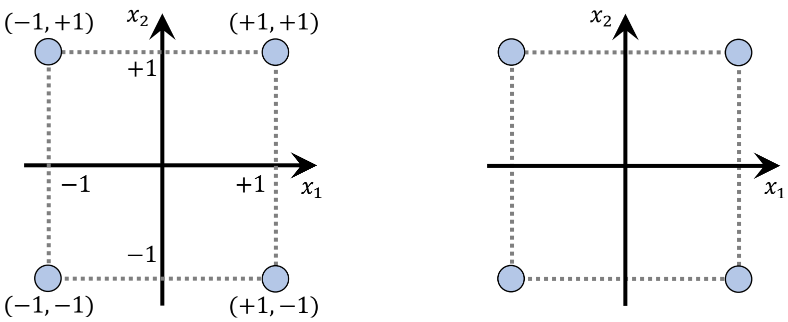
3. FACTORIAL TEST PLANS & ORTHOGONALITY

This section focusses on the demand of orthogonality for test plans, and how deviations from perfect orthogonality can be measured. To begin with, a short introduction to the considered test plans is given. Afterwards, orthogonality and a selection of potential non-orthogonality measures are introduced.

3.1 Full-factorial test plans

The scope of this paper are full-factorial test plans or designs, which are obtained by considering all possible parameter combinations on the chosen levels. For linear dependencies to be found, two levels per factor are sufficient. With k factors, this yields 2^k parameter combinations. Therefore, such test plans are also referred to as 2^k designs and are a common choice. For non-linear relationships, 3^k designs or central composite designs (CCDs) are used [10].
The levels in a 2^k design are chosen to be the minimum and maximum value of each parameter, in order to have the highest chance of detecting an influence (cf. Section 2.2). The parameters are usually normalized, yielding a unit free representation with "-1" being the lower level and "$+1$" being the upper level of each parameter, respectively. An exemplary 2^2 design – i.e., a 2^k design with two parameters – is depicted in Figure 2.

Figure 2: Graphic representation of a 2^2 test plan. The graphs show the parameter configurations (circles) of two parameters x_1 and x_2 on two load levels each. The load levels are normalized to -1 and $+1$. If this normalization is taken as given, the simplified graph on the right is equivalent.



An equivalent tabular representation, referred to as *design matrix*, is possible. Such a matrix can be used as foundation to connect the test planning to the evaluation. This connection is performed by extending the table with the columns for the additional quantities needed in multivariate linear regression (cf. Section 2.3). The constant term z_0 on level $+1$ is mandatory. In the case of a 2^2 design, the interaction term of the two inputs $z_3 = x_1 x_2$ has to be added as well. This yields the *matrix of independent variables* as given in Table 1, which coincides with the data matrix $\boldsymbol{Z}$ from eq. (5). It should be noted that it is common to replicate test runs by adding repeated rows to the matrices shown in Table 1.

Table 1: Design matrix with no. of test runs (left) and matrix of independent variables (right) for the considered 2^2 test design without repetitions. While the design matrix contains all relevant information for the conduction of the experiments, the matrix of independent variables contains the information relevant for the evaluation of the experiments.

Run	x_1 (z_1)	x_2 (z_2)
1	-1	-1
2	-1	$+1$
3	$+1$	-1
4	$+1$	$+1$

1 (z_0)	x_1 (z_1)	x_2 (z_2)	x_1x_2 (z_3)
$+1$	-1	-1	$+1$
$+1$	-1	$+1$	-1
$+1$	$+1$	-1	-1
$+1$	$+1$	$+1$	$+1$

3.2 The concept of orthogonality

Orthogonality in the sense used in DoE originates from the idea that if the true fundamental influences $x_1, \dots, x_m$ are considered, they are uncorrelated by definition. Possible correlation is important only when the interaction terms $x_i x_j$ are derived. This is why, as Figure 2 shows, the parameter axes are depicted orthogonally. This might seem trivial, yet only test plan orthogonality ensures that this theoretical orthogonality is not mitigated by the placement of experiments in the parameter space. In order to explain this principle in more detail, two extreme exemplary cases should be considered:

1. The heaviest violation of this idea would be a choice of experiments for which the parameters are linearly dependent (i.e., at least of them can be completely described by one or more other parameters). In this case, regression analysis is impossible, since there is no data which allows to decide whether an observed influence has to be accredited to one or another parameter. The equivalent mathematical reason can be derived from Section 2.3: the design matrix, X, and therefore also the data matrix Z would have no have full column rank. This makes the inversion of $\boldsymbol{Z}^T\boldsymbol{Z}$ impossible (cf. eq. (7)). Designing a test plan in a way that the chosen values of the input quantities are linearly dependent could therefore also be labeled as *perfect non-orthogonality*.
2. In the opposite – ideal – case, the chosen parameter values perfectly implement the theoretical orthogonality to the experiments conducted for every parameter. This situation might be referred to as *perfect orthogonality*. Then, OLS is possible and leads to minimum confidence bounds $\hat{\sigma}$ on the estimated coefficients [3].

From these two examples, it can be seen that it is not only necessary to avoid linear dependency (i.e., perfect non-orthogonality), but also beneficial to aim for perfect orthogonality, as Fisher originally recommended [1].

While perfect non-orthogonality is impossible to evaluate, orthogonality over the whole test plan is impossible to ensure. In any practical conduction, deviations from the planned parameter values will occur. These deviations might be neglected in the evaluation, but should actually be considered to get the most out of the data. It is therefore not enough to consider orthogonality as qualitative (i.e., yes or no) measure. In between these two extremes, a wide range of different degrees of non-orthogonality exists, which is highly relevant for conscious and efficient DoE.

3.3 Measuring non-orthogonality

For the detection of (non-)orthogonality with mathematical measures, the design matrix $\boldsymbol{X}$ and the data matrix $\boldsymbol{Z}$ need to be studied. Although the elements of $\boldsymbol{X}$ refer to the planned parameter values and the elements of $\boldsymbol{Z}$ refer to the real parameter values from the experiment, in the following, it will be assumed that both matrices contain the real parameter values to simplify the explanations. A non-exhaustive selection of methods and parameters to quantify non-orthogonality will be given and two parameters are chosen due to their characteristics for application in the parameter study. The initially considered quantities are:

- matrix orthogonality [2,3],
- column orthogonality [9],

- multicollinearity [11,12],
- variance inflation factor [13],
- reciprocal condition number [14],
- parameter correlation [9,11-13],
- the *A-optimality* measure A_2 [15,16], and
- the *D-optimality* measure D [15,16].

From these, some are difficult to apply due to analytic constraints, like e.g. matrix orthogonality. Others only offer a binary measure on whether a test plan is orthogonal or not, like e.g. column orthogonality. Another important aspect is whether the different measures are able to summarize the degree of non-orthogonality in a single number. This is especially helpful when analyzing systems of different size, or systems with many independent variables. Also, a single number is more qualified for practical application in future use cases. Last but not least, the codomain of the considered quantities is relevant. Each non-orthogonality measure is usually bounded on at least one side – that of maximum non-orthogonality or that of perfect orthogonality. Additionally, it can further simplify the interpretation of obtained values if the codomain is also bounded on the other end at a maximum degree of non-orthogonality.

According to these specifications, the two last mentioned quantities A_2 and D are chosen. They are both known for their use in the concepts of *A-optimality* and *D-optimality* and are calculated from the correlation matrix $\boldsymbol{R_Z}$ of the matrix of independent variables [9],

$$\boldsymbol{R_Z} = Corr\left(\boldsymbol{Z_{i,j>1}}\right). \tag{10}$$

Note that the elements of $\boldsymbol{R_Z}$ based on the first column of $\boldsymbol{Z}$, $\boldsymbol{z_0}$, are not defined. Therefore, the first column $\boldsymbol{z_0}$ is usually omitted when computing the correlation matrix based on eq. (29). With the elements r_{ij} of $\boldsymbol{R_Z}$, A_2 and D are obtained via

$$A_2 = \sum_{i<j}^{m} r_{ij}^2 , \tag{11}$$

$$D = \det(\boldsymbol{R_Z}) = \det\left(Corr(\boldsymbol{Z})\right) \in [0,1]. \tag{12}$$

While $A_2 = 0$ if and only if the design matrix is orthogonal, D is 1 in this case [15]. On the other end of the scale, D is bound to zero, while A_2 might increase to values higher than 1.

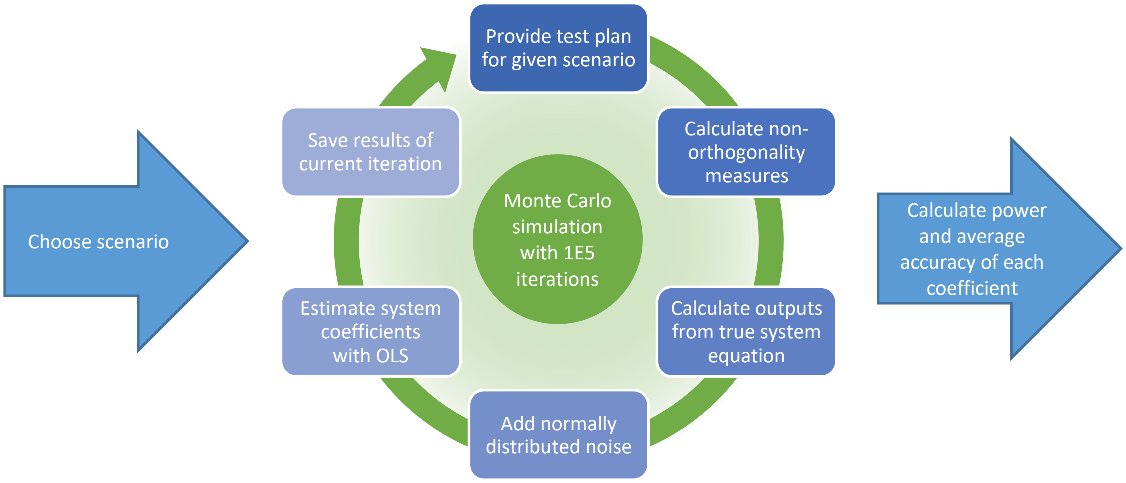
4. PARAMETER STUDY

To study the effects of non-orthogonality on test results, different scenarios of non-orthogonality in two exemplary systems are considered in a parameter study. The power and the accuracy of the model coefficients found per regression are calculated as results for each scenario. To additionally measure non-orthogonality, the A_2 and D as chosen in Section 3.3 are used. This parameter study serves as a preparatory screening measure to provide a profound basis for a future in-depth study. The overall simulation method, the considered non-orthogonality scenarios, the results for each system, and the summarized key findings are presented in the following.

4.1 Simulation procedure

The simulation procedure is shown graphically in Figure 3. For each scenario of non-orthogonality considered as well as for the perfectly orthogonal case, the manipulated non-orthogonal test plan is derived. The initially simulated orthogonal test plan is used as reference for comparison. Each single experiment is conducted 3 times for the two-dimensional and 2 times for the three-dimensional system, resulting in a total of 12 and 16 parameter combinations, respectively, for the orthogonal case. The results are then statistically evaluated using regression analysis as defined in Section 2.3. This process is repeated using the Monte Carlo method with 100,000 simulations. The mean and variance as well as the resulting power ($\alpha = 0.05$) of each model coefficient is computed. The results of the perfectly orthogonal as well as the different non-orthogonal test plans are compared.

Figure 3: Procedure of the conducted simulation. For the perfectly orthogonal case as well as for each considered case of non-orthogonality, a Monte Carlo simulation is performed. In each iteration, the virtual experiments according to the test plan for the given scenario are conducted. The results are evaluated using OLS. After 100,000 iterations, the power and the average accuracy of each coefficient is calculated.



4.2 Considered systems and cases of non-orthogonality

The two exemplary systems are linear systems with two and three input variables $x_1, x_2 (, x_3)$ and both with one output quantity y. The system definition according to eq. (1) therefore reduces to

$$y = c_0 + c_1 x_1 + c_2 x_2 + c_3 x_3 + c_{12} x_1 x_2 + c_{13} x_1 x_3 + c_{23} x_2 x_3 + \varepsilon, \tag{3}$$

$$y = c_0 + c_1 x_1 + c_2 x_2 + c_{12} x_1 x_2 + \varepsilon. \tag{4}$$

According to the dimension of the input, the systems will be referred to as two-dimensional and three-dimensional in the following. Their model coefficients, the error standard deviation σ and the derived maximum output y_{max} (without error ε) are given in Table 2. The coefficients for all inputs and interactions are set to equal levels. This might be an idealization, but is necessary to avoid an additional effect due to different parameter influences which interfere with the non-orthogonality.

Table 2: Coefficients, error standard deviation and maximum mean output for both systems

Inputs	c_0	c_1	c_2	c_3	c_{12}	c_{13}	c_{23}	c_{123}	σ	y_{max}
2	30	10	10	-	10	-	-	-	10	60
3	30	10	10	10	10	10	10	10	10	100

The considered sources and influences of non-orthogonality, which are examined consecutively, are:

(a) A systematic shift of a parameter away from its planned level. This serves to simulate a scenario in which a parameter level is not (or only with high effort) realizable with the system or was implemented wrongly by accident. Therefore, the fourth parameter combination with $x_2 = +1$ in the orthogonal case is changed to $x_2 = \{+0.9, 0, -0.9\}$ (denoted (a1), (a2) and (a3) in the following). These manipulations are shown in Figure 4 for the case of two input parameters.

(b) The removal of a single parameter combination. This represent the case that this particular experiment has to be omitted due to physical or timely constraints. Since the influence of each parameter is identical, it doesn't matter which parameter combination is omitted. This manipulation in presented in Figure 5 on the left.

(c) An additional repetition of a single parameter combination. This simulates the case that the experimenter wants to yield additional information by simply repeating an experiment. It should be noted that this is no non-orthogonality per definition, but adds imbalance to the test plan and serves to study this effect separately. As for the removal of a parameter combination, it doesn't matter which one is omitted. Figure 5 shows this scenario in the center.

(d) A random, normally distributed scattering σ_x is applied to the data matrix, moving every parameter combination slightly away from its perfectly orthogonal position. One possible test plan originating from this manipulation is shown in Figure 5 on the right. This scenario represents the ubiquitous uncertainty in the setting of the parameter levels, which might either be caused by a less cost-intensive test setup, a less time-intensive test process or physical constraints. The standard deviation of the scattering, $\sigma_x = 0.2$, is assumed to be known.

It is assumed that no more than one source of non-orthogonality appears simultaneously. Therefore, the mentioned scenarios are studied separately. This is permissible for screening test plans in a DoE context, as they serve to filter out the most relevant influences to prepare more thorough studies.

Figure 4: Graphic representation of the 2^2 test plan when the fourth parameter combination is shifted away from its orthogonal position $(+1, +1)$ along the x_2 axis towards $(+1, +0.9)$ (left), $(+1, 0)$ (center), or $(+1, -0.9)$ (right), to implement different degrees of non-orthogonality.

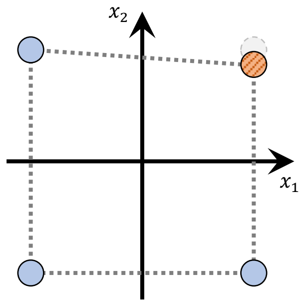
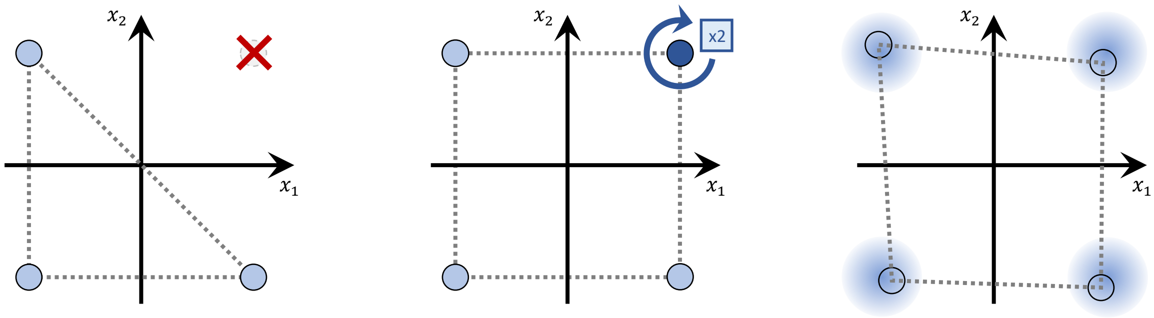


Figure 5: Graphic representation of the 2^2 test plan when the fourth parameter combination is omitted (left), repeated (center); or when all parameter combinations are superimposed with a normally distributed scattering around their orthogonal position (right).



4.3 Results

In the following, the results of both considered systems are presented and discussed in separate sections. The loss in test power and coefficient accuracy due to the non-orthogonal scenarios is presented and discussed in terms of practical application. In the end, the orthogonality measures A_2 and D are assessed.

System with two input variables

Table 4 presents the results of the parameter study with the system defined by eq. (4) in terms of the detectability (power) and the accuracy (estimation error) of the coefficients. In nearly every case, the systems general mean c_0 is guaranteed to be found. Only in the most extreme parameter shift (Scenario (a3)), or when a parameter combination is omitted (Scenario (b)) does the power of the intercept decrease. Examining the results in terms of the main influences x_1 and x_2, it can be seen that the gradual increase of non-orthogonality by shifting a parameter combination along the axis of x_2 decreases the power for the according coefficient, as one would expect (Scenarios (a1)-(a3)). Yet in addition, the detectability of the first component x_1 also decreases drastically, albeit with lower intensity. The same applies for the interaction x_{12}. This shows that non-orthogonality, even if it is limited to one parameter, corrupts the whole test plan.

The interaction itself is completely undetectable if one of the parameter combinations is omitted (Scenario (b)). This is the expected results, since any system with 3 fundamentally different parameter combinations – even if they are repeated – but 4 coefficients is underdetermined. Coefficient estimation results in this case are not reasonable; even worse, the power suggests that they are when it comes to the constant term c_0. Although computational workarounds for an underdetermined system exist (e.g. *ridge regression* [9]), this emphasizes how necessary a sufficient number of different parameter combinations is.

Scenario (c) shows that the studied test design can still be improved by solely repeating a single parameter combination in each repetition, leading to 15 instead of 12 experiments and a power of 92 % instead of 86 % per parameter. This indicates that the positive effect of the increased number of experiments is greater than the negative effect of the non-orthogonality added to the test plan.

Last but not least, a normally distributed offset added to each parameter level (Scenario (d)) seems not to deteriorate the parameter detectability or accuracy. This is an interesting result, as in real-world applications, the experimenter will not be able to set the parameters to the exact levels defined by the theoretical test design. They will rather be scattered near their optimal position. The variance of this scattering depends on the system, the experimenter and especially the setup used. The last point is important when considering the cost of an experiment, where the experimenter might have to decide between more expensive (but more exact) or less expensive (and also less exact) hardware. In this mindset, the result of the last scenario suggests that deviations at the given level are negligible, indicating that this kind of deviation should be assessed in further studies.

While the coefficient estimation accuracy more or less limits to a high and a low level, the power is more continuous. The power of the two independent variables and their interaction can be linearly predicted by A_2 with a R^2 between 93 % and 98 %. D shows even higher linear correlation with R^2 between 95 % and 99 %. This indicates that both quantities are able to measure non-orthogonality in the given case. However, further studies are needed to assess how A_2 and D behave for the different deviations from orthogonality to obtain a comprehensive relationship.

Table 4: Results of the test of the two-dimensional system with 3 repetitions. The first row shows the orthogonal reference case. The following rows show the scenarios described in Section 4.2: a parameter shift in 3 variations (a1-a3), the omission (b) and the repetition (c) of a parameter combination as well as a random scattering of parameters around their orthogonal levels (d). The columns show the non-orthogonality measure A_2 and D, the power $(1-\beta)$ and the estimates mean deviation from the real values for each parameter.

Scenario	Non-orthogonality		Power $(1-\beta)$ [%]				Coefficient estimation error [%]			
	A_2	D	c_0	c_1	c_2	c_{12}	c_0	c_1	c_2	c_{12}
orthogonal	0	1	100.0	85.9	85.9	85.8	0	-0.1	0.1	0.2
(a1)	0.004	0.996	100.0	85.7	84.0	83.8	0	0	0	0
(a2)	0.388	0.529	100.0	69.8	48.6	48.6	0	0.1	0.2	0
(a3)	0.762	0.005	10.4	5.6	5.5	5.5	-0.4	-1.0	-1.2	-1.2
(b)	0.750	0	97.4	4.2	4.2	0	-33.4	-100.2	-100.2	-100.0
(c)	0.083	0.926	100.0	92.0	91.8	92.1	0	-0.1	-0.1	0.1
(d)	0.038	0.963	100.0	85.6	85.5	86.1	-0.3	-0.1	-0.1	-0.1

System with three input variables

The power of the coefficients in the three-dimensional system defined in eq. (3) are shown in Table 5. The estimation errors are equivalent to those in the two-dimensional system and are therefore omitted here. The results regarding the general mean c_0 are also similar to the previous case: Its power decreases drastically in the most radical shift of the second component in one parameter combination (Scenario (a3)). Yet, the power of c_0 is about twice as high as for the two-dimensional system (19.6 % instead of 10.4 %). This is probably due to the greater total number of experiments in the bigger system (16 instead of 12). Assessing the three variants of Scenario (a) together, it seems that the findings from the two-dimensional system are repeated: Not only the power of the coefficients connected to x_2 (i.e., c_2, c_{12}, c_{23}, c_{123}; shaded grey in Table 5) decreases gradually, but also the power of every other parameter. Yet, a new insight arises from the three-dimensional system, as the power for the coefficients depending on x_2 decreases faster than that of the other coefficients. This can best be seen in Scenario (a2), where the power of the first mentioned is approximately 76 %, while that of the latter is 88 %.

The results of Scenario (b) mirror what was already seen in the two-dimensional system. The repetition of one of the fundamental parameter combinations, Scenario (c), increases the power of all parameters from approximately 94 % to 96 %. This means that the positive effect of an increased number of experiments is small here. It has to be assessed in further studies if this power increase could be greater if a less non-orthogonal arrangement of the repeated parameter combination is beneficial.

The last scenario, where a normally distributed noise has been applied to the parameter levels, shows a slightly decreased power for all parameters (93 % instead of 94 %). This effect is clearer here than in the case of the two-dimensional system. However, the power loss is still small. If the cost or duration of an experiment depends on how exact the parameter levels are reached, an appropriate amount of such scattering may be permitted to make the test more efficient. It has to be further assessed how different amounts of noise applied to the parameter levels affect the resulting test power.

When regressing the power of the different coefficients on either A_2 or D, the results show a R^2 between 76 % and 81 % in both cases, which is less than in the two-dimensional system. Even more than for the two-dimensional system, it has to be studied if A_2 and D correlate better with non-orthogonality if only a single source of it is present.

Table 5: Results of the test of the three-dimensional system with 2 repetitions. The first row shows the orthogonal reference case. The following rows show the scenarios described in Section 4.2: a parameter shift in 3 variations (a1-a3), the omission (b) and the repetition (c) of a parameter combination as well as a random scattering of parameters around their orthogonal levels (d). The columns show the non-orthogonality measures A_2 and D and the power $(1-\beta)$ for each parameter. The columns of coefficients connected to x_2 are shaded grey.

Scenario	Non-orthogonality		Power $(1-\beta)$ [%]							
	A_2	D	c_0	c_1	c_2	c_3	c_{12}	c_{13}	c_{23}	c_{123}
orthogonal	0	1	100.0	93.7	93.7	93.7	93.7	93.7	93.7	93.7
(a1)	0.006	0.994	100.0	93.6	93.2	93.7	93.0	93.7	93.1	93.0
(a2)	0.379	0.458	100.0	87.9	75.5	87.8	75.5	88.0	75.5	75.7
(a3)	0.781	0.004	19.6	6.5	6.4	6.5	6.3	6.5	6.4	6.4
(b)	0.583	0	99.7	4.4	4.5	4.5	4.4	4.4	4.5	0.0
(c)	0.210	0.850	100.0	95.9	96.1	96.0	96.0	96.0	96.0	96.1
(d)	0.255	0.774	100.0	92.9	92.8	92.9	92.9	92.6	92.8	92.6

Key findings

The simulation results can be summarized in the following key findings:

- the estimation accuracy showed no more than 1.2 % error in the different non-orthogonality scenarios, as long as the test design is consistent;
- the estimation error was not found to be related to the degree of non-orthogonality;
- the power of the coefficients was found to be related to the degree of non-orthogonality
- small and medium offsets as well as a normally distributed scattering of parameter levels yield surprisingly small power losses;
- additional non-orthogonal replications show less benefit than expected, emphasizing the practical relevance of test plan orthogonality;
- the A_2 and D values, two optimality measures, were found to be linearly related to the power and therefore qualify as non-orthogonality measures.

5. CONCLUSION

In this paper, the basic principles of hypothesis testing, linear regression analysis and DoE as well as their relation to one another have been explained. The requirement of orthogonality in test design has been illustrated. Different methods and quantities to measure deviations from perfectly orthogonal test plans – which will always be present in application – have been presented. Test power – i.e., detectability of existing influences – and estimation accuracy have been defined as evaluation criteria to assess the quality of a test design. Both quantities have been examined in a screening parameter study which implements common sources of non-orthogonality in test plans. These sources have been explained and realized in two systems with two and three input variables, respectively. The simulation has been conducted using the Monte Carlo method. The results have been presented, discussed and compared to two different non-orthogonality measure.

Although the systems considered in this paper were quite basic, they were already able to help assess the fundamental meaning of test plan (non-)orthogonality. Also, first hints towards more cost- or time-efficient test design in practical applications were found. As a follow-up, an in-depth study with several extensions is suggested. For example, different coefficients for the considered parameters as well as for their interactions should be considered. More generalized insights could be found analyzing the p-value instead of the power. Also, more variations of the considered non-orthogonal scenarios are suggested. Such further examinations should be considered in the pursuit of efficient test plans.

References

[1] R. A. Fisher, "*The Design of Experiments*", Oliver & Boyd, 1935, Edinburgh.

[2] D. C. Montgomery, "*Design and Analysis of Experiments*" (8th ed.), John Wiley & Sons, 2013, Hoboken, NJ.

[3] W. Kleppmann, "*Versuchsplanung. Produkte und Prozesse optimieren*" [Experimental design: Optimizing Products and Processes], Hanser, 2013, Munich.

[4] K. Siebertz, D. van Bebber and T. Hochkirchen, "*Statistische Versuchsplanung. Design of Experiments (DoE)*" (2nd ed.), Springer, 2017, Berlin.

[5] G. E. P. Behnken, W. G. Hunter and S. Lang, "*Statistics for Experimenters*" (2nd ed.), John Wiley & Sons, 2005, Hoboken, NJ.

[6] G. Taguchi and S. Konishi, "*Taguchi Methods Orthogonal Arrays and Linear Graphs, Tools for Quality Engineering*", American Supplier Institute, 1987, Dearborn, MI.

[7] J. F. Box, "*R. A. Fisher and the Design of Experiments, 1922-1926*", The American Statistician, vol. 34, pp. 1-7, (1980).

[8] J. M. Wooldridge, "*Introductory Econometrics. A Modern Approach*" (5th ed.), South-Western, 2013, Mason, OH.

[9] A. C. Rencher and G. B. Schaalje, "*Linear models in statistics*" (2nd ed.), John Wiley & Sons, 2008, Hoboken, NJ.

[10] M. Arndt, P. Mell and M. Dazer, "*Generic effects of deviations from test design orthogonality on test power and regression modelling of Central-Composite Designs*" [Manuscript submitted for publication], PSAM 16, 2022, Honolulu, Hawaii.

[11] F. Öztürk and F. Akdeniz: "*Ill-conditioning and multicollinearity*", Linear Algebra and its Applications, vol. 321, pp. 295-305 (2000)

[12] D. A. Belsley and R. W. Oldford: "*The general problem of ill conditioning and its role in statistical analysis*", Computational Statistics & Data Analysis, vol. 4, pp. 103-120 (1986)

[13] R. M. O'Brien: "*A Caution Regarding Rules of Thumb for Variance Inflation Factors*", Qual Quant (Quality & Quantity), vol. 41, pp. 673-690 (2007)

[14] F. W. J. Olver, D. W. Lozier, R. F. Boisvert and C. W. Clark, "*NIST Handbook of Mathematical Functions*", Cambridge University Press, 2010, New York City

[15] H. Xu: "*An Algorithm for Constructing Orthogonal and Nearly-Orthogonal Arrays with Mixed Levels and Small Runs*", Technometrics, vol. 44, pp. 356-368 (2003)

[16] V. V. Fedorov, "*Theory of Optimal Experiments*", Academic Press, 1972, New York City.