ACTA ET COMMENTATIONES UNIVERSITATIS TARTUENSIS DE MATHEMATICA
Volume 16, Number 2, 2012
Available online at www.math.ut.ee/acta/

On supersaturated experimental design

Sven Ahlinder and Ivar Gustafsson

Abstract. In this article, optimization of a black box using designs of experiments is studied. There are two fundamental different branches in optimization using Design of Experiments; higher order Taylor expansions, i.e., response surface methodology, and underdetermined experimental designs, i.e., supersaturated designs. In this article, supersaturated designs are studied.

Optimization using design of experiments is compared with direct search. A system of linear equations has to be solved, $Ax = b$, where A is the design matrix and b is a vector of experimental results. In supersaturated design, the solution $\hat{x}$ to this system is not in general the gradient of the system but has a certain angle to it. This angle between the correct gradient, x, and the one found by supersaturated design of a linear approximation of a black box function, $\hat{x}$, is investigated.

1. Introduction

In the business of construction (for instance of a car) the purpose is often to fulfill demands by changing parameters of the studied system. The corresponding discipline in statistics is called design of experiments (DoE). Usually in DoE, the number of observations is larger than the number of variables, and we have so called overdetermined designs. This gives a possibility to estimate statistical properties of the coefficients that describe the influence of the variables such as standard deviation and possibility to perform an F-test on the coefficients.

Already in 1935, R. A. Fisher published a book "The Design of Experiments" [6] describing how to design experimental series in order to be able to estimate coefficients in linear polynomials in several variables. These polynomials then served as models to be able to maximize the yield of crops.

In vehicle industry, DoE is used in studying several problems such as fuel consumption, emissions, noise, durability, handling, welding, painting,

Received July 29, 2010.
2010 *Mathematics Subject Classification.* 15A06, 14A09, 62K15.
Key words and phrases. Supersaturated design, singular value decomposition.

vibration, air resistance, cooling, heating, etc. The main problem is that a modern vehicle is specified by say million parameters and each experimental vehicle costs perhaps $1.000.000. The companies can not afford to make a million experiments. This phenomenon has been identified already in the 1950ies and the solution to it was found by application of supersaturated design. We will define this kind of designs below.

Suppose we have a physical system, for instance a car. Suppose also that there is an influence on fuel consumption from the different parameters of the car, such as shape of body, engine, wheels, and even different tyre pressure. Suppose the influence of the different parameters on fuel consumption has certain values. Then those values could be gathered in a vector x, which is the true description of the system with (say) n parameters. We can alter the parameters and see if the fuel consumption is increasing or decreasing. The (say) m parameter settings of different cars could be gathered in row vectors $a_j,\ j = 1, ..., m$, and the whole experimental series would form a matrix A, with the different vectors. For instance, we let the series of m different fuel consumptions, from different experimental cars, be the vector b, i.e., we operate on the system description x with our experimental design A and get a result vector b, that is $Ax = b$, where $A \in \mathbb{R}^{m\times n},\ x \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$. Normally A is overdetermined, i.e., $m > n$, which means that the number of experiments m is larger than the number of parameters n.

A supersaturated experimental design is a plan to create an experimental series with fewer observations than unknown parameters, $m < n$, i.e., the system $Ax = b$ is underdetermined. It is well known that an underdetermined system of linear equations could be solved by a generalized inverse based on singular value decomposition (SVD), giving the smallest norm solution.

At Volvo, supersaturated design (SSD) optimizations have been tested on mathematical functions [11]. Some theoretical functions and a genuine engine model were tested. In all cases about 100 variables were used and 50 evaluations were performed. The 50 evaluations were divided into 3 steps with about 15 evaluations at each step. For each step, centre of the investigation was moved in the direction of steepest ascent. In all cases the function value was improved for each step. In fact, 2000 random experiments did not beat 50 SSD experiments, which indicates that the method has potential.

In normal, overdetermined (nonsaturated) designs, the purpose is to approximately replicate the measured results, given by the vector b. This results in methods for estimating predictions, $\hat{b}$, which are given by experimental series, say A, that have orthogonal columns since this gives the possibility to estimate the elements in x independently of each other. According to Fisher, it is useful to add some extra rows to A, without destroying the

orthogonality, to make the investigation less sensitive to measurement errors. In overdetermined designs, the system of equations can be solved with the standard method of least squares, i.e., $\hat{x}$ is the solution to the problem $min_x \|Ax - b\|_2$.

In this paper a new method is introduced for studying supersaturated designs. Here it is assumed that the first order Taylor series describing the black box is known. The first order Taylor series is the sum of all parameters each multiplied by a coefficient. The coefficients are the elements of x. The coefficients give an ascent direction as our search direction in an optimization step. When the ascent direction step has been taken, a new estimation of the coefficients can be made in an iterative way. The method is focused on how to estimate the coefficients, the elements of a vector $\hat{x}$, using the experiments A on the correct vector x. For an overdetermined (nonsaturated) A, $\hat{x} = x$ and this is not a problem, but for a supersaturated (underdetermined) A this is vital since, in general, then $\hat{x}$ is not equal to x.

If x lies in the row space of A, then the experimental series will be a success, since then $\hat{x} = x$, but if x lies in the null space of A, it will be a complete failure. In all practical cases, x has components in both, the row space and the null space. The best we can do is to take an A with a dimension of the row space as large as possible compared with the null space. This means that the rank, or the number of independent rows of A, should be as large as possible. This leads to the fact that in supersaturated designs, A should have linearly independent rows, while in nonsaturated designs, A should have linearly independent (or even orthogonal) columns.

One of the main results of this paper is that the total gain by using supersaturated design correctly is expected to be $\sqrt{m \cdot n}$. This means that investigating four times more parameters with the same amount of tests gives doubled information.

The main conclusions of this article are the following.

- Designing an experimental series using minimum norm solution gives different designs than designing it for variable selection.
- The rank of the design matrix A is important for the information gain in the system, not the angle between the columns of the design matrix A.
- If A is $m \times n$, $m < n$, x is a normally distributed vector, and $\hat{x} = A^+ Ax$, where A^+ is the Moore-Penrose pseudo inverse, then the mean

$$E(|\hat{x}^T x|/(\|\hat{x}\|\|x\|)) \to \sqrt{m/n},$$

when $m \to \infty$ and the quotient m/n is kept constant. Here $\|\cdot\| = \|\cdot\|_2$ is the Euclidean norm of a vector.
- The expected information of a supersaturated design is $\sqrt{m \cdot n}$.

2. Basic background for design of experiments

2.1. Design of experiments and optimization. A gradient optimizer calculates an approximate gradient of a system, for instance a function, at some point p_0 and goes in this direction, which is an approximation of the steepest ascent when maximizing. The gradient could be approximated by taking a small step $\delta > 0$ from p_0 in each variable one at a time and calculate the influence on the function for each variable. In DoE terminology, all experiments of one gradient evaluation form a matrix A that is diagonal with diagonal elements equal to δ, i.e., $A = \delta I$. Since A is invertible, $x = A^{-1}b = \delta^{-1}b$ which is the same as dividing all elements in b by δ. This could be seen as numerical differentiation. Then a step, normally much larger than δ, is taken in the direction of the gradient. This larger step is normally repeated several times until an optimum is reached. Then the system (function) is evaluated again and a new direction is computed.

In general, DoE basicly works in the same way as a gradient optimizer. One calculates an approximation to the gradient by $\hat{x} = A^{+}b$ and moves in the direction $\hat{x}$, where A^{+} is a suitable inverse. The four major differences regarding general DoE, compared to gradient opimization, are the following.

- Normally there is only one, reasonably large, step for each system evaluation.
- DoE takes a large step δ when evaluating the gradient which gives the system a solid kick [2]. The choice of step length is crucial. The longer step is taken when investigating the gradient, the more exactly the influence of the variable is estimated. On the other hand, if the step is too long, we can miss the curvature. The gradient solver investigates only a small surrounding of the starting point and then takes a larger step. When taking the larger step you get into an area that is not investigated. One compromise is to let the step length in a gradient investigation be as long as the planned step length in the improvement step.
- In DoE, A is not just I, but it has balanced columns. This means that in each column of A there are equally many steps in the positive direction, plus ones, as in the negative direction, minus ones. This leads to the fact that only certain combinations of rows and columns are orthogonal and those are called experimental designs. One purpose of balanced designs is that the influence of the experimental error is divided by the number of rows [2]. We note that in this paper we only study designs for two-levels factors.
- A is mostly overdetermined in DoE, i.e., A has more rows than columns, so the influence of errors is decreased. Overdetermined systems of equations are often solved by the method of least squares. Recall that A should be chosen with regard to properties as cost of experiments and orthogonality.

This means that in DoE one decides in advance which system of equations should be solved by deciding which experiments should be performed on the system.

2.2. Optimization with supersaturated designs. A supersaturated design (SSD) is an experimental series with fewer observations than variables. The derived underdetermined system of equations can be solved by the Moore–Penrose pseudo inverse, $\hat{x} = A^+ b$. The Moore–Penrose pseudo inverse is defined by $A^+ = A^T(AA^T)^{-1}$ if A has full rank, i.e., if $rank(A) = m$, and by a singular value decomposition of A (SVD) in case $rank(A) < m$. For a short introduction to SVD, see Section 3.2. This pseudo inverse gives the minimum norm solution of $Ax = b$, i.e., the solution with the smallest (Euclidean) norm $\|\hat{x}\|_2$ [5]. Note that the scaling of the problem is vital.

Except for using pseudo inverse solution, optimization with SSD does not differ from ordinary (overdetermined) DoE. The solution x, the correct gradient or steepest descent, is estimated by $\hat{x}$ and a step along this direction is taken with a reasonable step length [11].

The traditional methods for generating supersaturated designs, see for instance [8], mainly focus on low pairwise correlation between the columns. This is called $E(s^2)$ technique, which means that the maximal correlation of the columns of A should be as small as possible. This is good in the typical case of (overdetermined) DoE, since $\hat{b}$ not neccessarily equals b. However, SSD is underdetermined and thus atypical in DoE as will be shown later, since $\hat{b} = b$. The important concern in SSD is that A has linearly independent rows, not columns, i.e., A should have full rank or almost full rank.

3. Theory for supersaturated designs

3.1. Solving the system of equations. In order to estimate the coefficients x we have to solve a system of linear equations to get an approximation $\hat{x}$. In the case of supersaturated designs the system is underdetermined and we will use a pseudo inverse based on SVD in order to compute the minimal norm solution to the system, see [5]. However, computing SVD of a large system is expensive. This is not a severe problem in our applications, since supersaturated designs are aimed for expensive experiments which dominate the total cost anyway.

3.2. Singular value decomposition (SVD). Any matrix $A \in \mathbb{R}^{m\times n}$ can be factorized into a product of three matrices $A = U\Sigma V^T$, where $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ are orthogonal and $\Sigma \in \mathbb{R}^{m\times n}$ is nonnegative and quasi-diagonal, i.e., all elements but σ_{ii}, $i = 1, \ldots, n$, are zeroes, see [5]. The diagonal elements σ_{ii}, $i = 1, \ldots, n$, of Σ are called singular values of A. Since the inverse of an orthogonal matrix is its transpose, we may define a generalized inverse by $A^+ = V\Sigma^+ U^T$, where Σ^+ is the transpose of Σ

with all nonzero elements replaced by their inverted values, i.e., $\Sigma^+ \in \mathbb{R}^{n\times m}$ is a quasi-diagonal matrix with diagonal elements $\frac{1}{\sigma_{ii}}$, $i = 1, \ldots, n$. Here we assume that A has full rank. It follows that $\hat{x} = A^+b$ is the unique solution (except from rounding errors) to the system $Ax = b$ if $m = n$ with $rank(A) = m = n$ (i.e., the full rank case with $A^+ = A^{-1}$), least square solution to $Ax = b$ if $m > n$ and minimum norm solution to $Ax = b$ if $m < n$. In the latter case A^+ is the Moore–Penrose pseudo inverse. We note that there exists a compact form for SVD, $A = U_1\Sigma_r V_1^T$, where $\Sigma_r \in \mathbb{R}^{r\times r}$, $r = rank(A)$, and $U_1 \in \mathbb{R}^{m\times r}$, $V_1^T \in \mathbb{R}^{r\times n}$. This form is used when A is rank defective, i.e., when $rank(A) < min(n, m)$.

3.3. Estimates. In this paper we assume x as a property of the object studied. The estimated gradient is called $\hat{x}$. For supersaturated designs, x and $\hat{x}$ are not necessary the same. The vector x is regarded as the correct gradient, the first order Taylor expansion of the object studied which in this paper is assumed to be known. The matrix $A \in \mathbb{R}^{m\times n}$ includes the experiments conducted on the object. Let b be the results of the experiments. If we want to estimate the first order Taylor expansion of the behaviour of the object, then we compute $\hat{x} = A^+b$, where A^+ is the Moore–Penrose pseudo inverse. If we then want to check if the object behaves linearly, i.e., if $Ax = b$, we may calculate $\hat{b} = A\hat{x}$ and see if $\hat{b} = b$. The total calculation may be summarized as $\hat{b} = AA^+Ax$.

For $m > n$, $A^+ = (A^TA)^{-1}A^T$ is a left inverse so $A^+A = I$. This means that $\hat{x} = A^+Ax = Ix = x$ and $\hat{b} = A\hat{x} = AA^+b$, which is not b if the system is not linear or b has noise in the measurements, so in least squares you get correct coefficients but (in general) wrong estimates, i.e., $\hat{b} \neq b$. The matrix AA^+ is very well studied and is called *the hat matrix*, $H = AA^+$ [12]. The matrix A^+A is not very well studied. Some properties of it are included in this paper. We call $A^+A = C$ *the cap matrix* and it is as important for underdetermined estimation as H is for overdetermined estimation.

For $m < n$ and A having full rank, $A^+ = A^T(AA^T)^{-1}$ is a right inverse so $AA^+ = I$. This means that $\hat{b} = AA^+Ax = IAx = Ib = b$, which means that the estimates given by the model is exact at the measured points. Hence, the residuals are zero and give no information. Note that $C = A^+A$ is not necessarily equal to I. This means that $\hat{x} = A^+Ax = Cx$ is not necessarily equal to x and the model gets (in general) wrong coefficients, i.e., $\hat{x} \neq x$, although $A\hat{x} = Ax$. Still it is possible to estimate an approximative angle, v, between $\hat{x}$ and x. The expected value of $\cos v$, where v is this angle, is $\sqrt{m/n}$ for large values of m and n, see Section 4.

If we have a new observation vector a_0, which is a linear combination of the rows of A, i.e., $a_0 \in Row(A)$, the row space of A, then $a_0\hat{x} = a_0x$. Hence, the quality of A dependends on $rank(A) = dim(Row(A))$ and not on the correlations of the columns as it is indicated in [8].

3.4. Some properties regarding pseudo inverses and estimates. In the following we let the vector norm be the Euclidean 2-norm, and the matrix norm be the corresponding norm, i.e., $\|A\|_2 = sup_{x\neq 0}\frac{\|Ax\|_2}{\|x\|_2}$. It is well known that $0 \leq \|Ax\| \leq \|A\|\|x\|$, $\|Ax\| = 0$ if $x \in Nul(A)$, and $\|Ax\| = \|A\|\|x\|$ if x is parallel to the column in V^T (the factor in the SVD of A, see Section 3.2) that corresponds to the largest singular value of A [5].

We consider the general, possibly not full rank, case for supersaturated designs, so $A^+ = V\Sigma^+U^T$, the Moore–Penrose pseudo inverse from Section 3.2. Let $C = A^+A = V\Sigma^+U^TU\Sigma V^T = V\Sigma^+\Sigma V^T$. The matrix C is symmetric since $\Sigma^+\Sigma$ is diagonal, see Section 3.2. Obviously, $CC = C$, i.e., C is idempotent, so the eigenvalues of C are zeroes or ones. It follows that $\|C\| = \sqrt{\lambda_{max}(C^2)} = 1$, where λ_{max} is the largest eigenvalue. Since $x \in Nul(A)$ implies $x \in Nul(C)$, we also get $\|Cx\| = 0$ if $x \in Nul(A)$.

Consider now the compact SVD of A, $A = U_1\Sigma_rV_1^T$, see Section 3.2. Then $A^T = V_1\Sigma_rU_1^T$ and the column spaces of A^T and V_1 are the same, i.e., $Col(A^T) = Col(V_1)$. Further, $C = A^+A = V_1\Sigma_r^{-1}U_1^TU_1\Sigma_rV_1^T = V_1V_1^T$ is then the orthogonal projection on $Col(A^T) = Row(A)$, the row space of A. In particular, if x belongs to $Row(A)$, then $\hat{x} = Cx = x$. From the fact that $C = V_1V_1^T$, with V_1 having r linearly independent columns, it follows that C has r eigenvalues equal to 1 and $n - r$ eigenvalues equal to 0. Thus $r = rank(C)$ is the sum of the eigenvalues of C, i.e., $r = trace(C)$, the sum of the diagonal elements of C.

3.5. Volvo data and Pareto principle. In 1906, Vilfredo Pareto observed that 80% of the property in Italy was owned by 20% of the population. In 1941, Dr. Joseph Moses Juran expanded the observation to "80% of your sales come from 20% of your clients". This is also known as the 80–20 rule or the vital few and the trivial many [9].

In Figure 1 we present a so-called Pareto diagram. The analytic function behind the graph for the normal probability distribution (the thick line), and also its inverse function, are given in Lemma 3.1 below. As we will see in the Lemma, the inverse function is much easier to use when drawing the function in a diagram.

We can see that the data gathered here is more close to a normal probability distribution than the Pareto principle. Therefore we are going to use normally distributed x in the further reasoning.

3.6. Creating lean designs. It was made clear in Sections 3.3 and 3.4 that $\hat{x} = Cx$ is the orthogonal projection of x on $Row(A)$ and that the dimension of this space should be as large as possible. The most common idea in supersaturated designs so far has been having correlations as small as possible between the columns of A using the $E(s^2)$-technique, see for instance

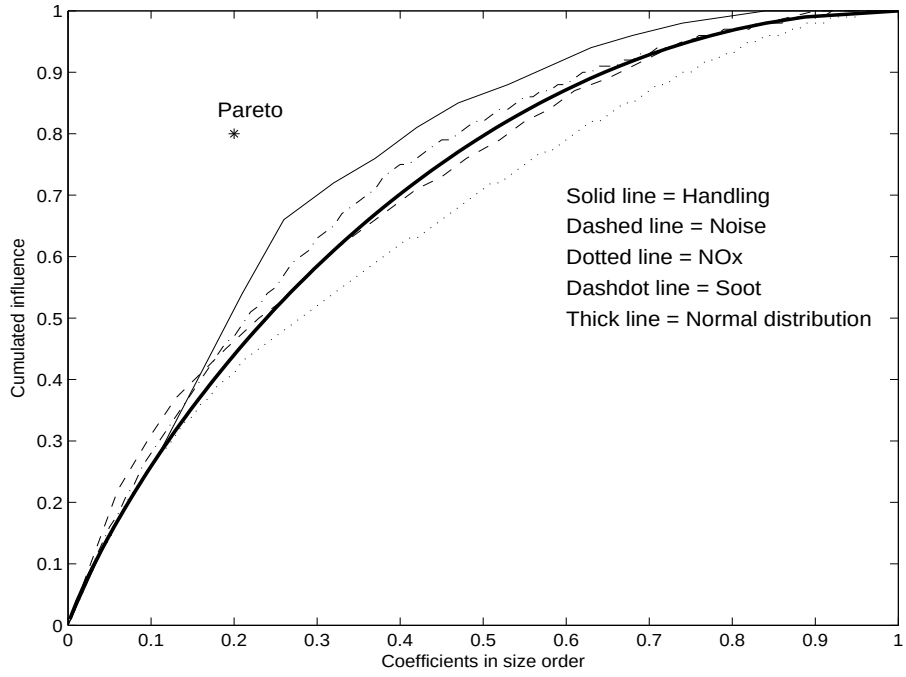

FIGURE 1. The Pareto principle versus normal distribution. The scaled accumulated coefficients of a number of massive industrial investigations (regarding handling, noise, NOx and soot) have been plotted in size order. The star is the point that the Pareto principle indicates. The thick line is the normal probability distribution, i.e., for all coefficients being $N(0,1)$.

[8], but from our analysis it becomes clear that it is the independency of the rows that is important, not the independency of the columns.

Orthogonal rows can easily be created by transposing classical designs. In Appendix 1 we present a MATLAB program for generating designs of this kind, so called lean designs. The special kind given by this algorithm is denoted AG-designs in [7], thereby refering to the authors of this paper (Ahlinder and Gustafsson).

Another way is to concatenate a suitable number of unit matrices I so that $A = [I, I, I, ..., I]$. This is equivalent to lumping variables together so that the first row consists of ones for the probably most influential variables and zeroes for the other, then the next row is ones for some of the less probable variables and zeroes for the others and so on.

The thick line in the Pareto diagram, Figure 1, represents the normal probability distribution. The line is generated by reordering the elements of

x, where x_i is $N(0,1)$, in decreasing order of magnitude, defining a vector y by $y_i = \sum_{k=1}^{i} |x_k|$ and finally is plotted y. In order to find an analytical function behind this graph we consider

$$F(x) = \int_z^\infty te^{-t^2/2}dt, \tag{1}$$

where x and z are related by the following equality

$$x = \frac{2}{\sqrt{2\pi}} \int_z^\infty e^{-t^2/2}dt.$$

Here the factor t in the integrand in (1) represents the size of an element and the factor $e^{-t^2/2}$ represents the corresponding probability density.

In the following Lemma we derive an implicit formula for the function $F(x)$ in (1), involving the normal distribution function. For the inverse of F we obtain an explicit formula.

Lemma 3.1. *The function F in* (1) *is given by the implicit formula*

$$F(x) = e^{-[\Phi^{-1}(\frac{2-x}{2})]^2/2},$$

where Φ is the normal distribution function, and its inverse function is given by

$$F^{-1}(y) = \frac{2}{\sqrt{2\pi}} \int_{\sqrt{-2\ln y}}^\infty e^{-t^2/2}dt. \tag{2}$$

Proof. The values x and z in (1) should be related by $x = \frac{2}{\sqrt{2\pi}} \int_z^\infty e^{-t^2/2}dt$, where the integrand represents the probability density and $\frac{2}{\sqrt{2\pi}}$ is the required normalizing factor. By the normal distribution function

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2}dt = 0.5 + \frac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2/2}dt$$

and since

$$x = \frac{2}{\sqrt{2\pi}} \int_z^\infty e^{-t^2/2}dt = 1 - \frac{2}{\sqrt{2\pi}} \int_0^z e^{-t^2/2}dt,$$

we get

$$2\Phi(z) = 1 + \frac{2}{\sqrt{2\pi}} \int_0^z e^{-t^2/2}dt = 2 - x \Leftrightarrow z = \Phi^{-1}(\frac{2-x}{2}).$$

We also note that $\int_z^\infty te^{-t^2/2}dt = e^{-z^2/2}$ so we may write the desired function

$$F(x) = e^{-[\Phi^{-1}(\frac{2-x}{2})]^2/2}.$$

For the inverse of this function we obtain a simpler, explicit formula (2). Let

$$y = \int_z^\infty te^{-t^2/2}dt, \ 0 \le y \le 1.$$

Then $y = e^{-z^2/2}$, $\ln y = -z^2/2$ and $z = \sqrt{-2\ln y}$. Further

$$x = \frac{2}{\sqrt{2\pi}} \int_z^\infty e^{-t^2/2} dt = \frac{2}{\sqrt{2\pi}} \int_{\sqrt{-2\ln y}}^\infty e^{-t^2/2} dt$$

so the inverse function of F is defined by

$$F^{-1}(y) = \frac{2}{\sqrt{2\pi}} \int_{\sqrt{-2\ln y}}^\infty e^{-t^2/2} dt.$$

□

4. Results from the present study

4.1. Information using supersaturated designs. In [3], designs are discussed that maximize the probability to identify active factors. This discussion leads to orthogonal columns so that the authors can make variable selection. In this paper, we solve the system by using the Moore–Penrose pseudo inverse giving the minimum norm solution. It turns out that not the columns should be independent, but the rows. It is the rows that carry the information further into the coefficients of the solution.

Recall that $C = A^+A$, where $A \in \mathbb{R}^{m\times n}$ with $m < n$, is a projection matrix that projects x on $Row(A)$. A measure of the efficiency of a design A is $\cos v$, where v is the angle between x and $\hat{x} = Cx$. If, for instance, $\cos v$ is 0.5, the gain is half the distance we are going. Let s be the saturation, i.e., $s = \frac{n}{rank(A)}$. One of the main results of the present study is that $\cos v$ goes to $\frac{1}{\sqrt{s}}$ when n goes to ∞. This statement is proved in Theorem 4.1 below. At first we state and prove a lemma. We consider vectors $x \in \mathbb{R}^n$ with components being independent normally distributed random variables with expectation 0 and variance 1, i.e., $x_j \sim N(0,1)$, $j = 1, \ldots, n$. We use the notation $x \sim N(0, I_n)$ for such vectors. It is just for simplicity we consider the variance equal to 1. Similar results can be given for a general variance.

Lemma 4.1. *Let $x \in \mathbb{R}^n$ be such that $x \sim N(0, I_n)$. Then the expected value of the norm of x is*

$$E(\|x\|) = \begin{cases} \frac{k!2^{k+1}}{\sqrt{2\pi}(2k-1)!!}, & n = 2k+1 \\ \frac{\sqrt{2\pi}(2k-1)!!}{2(2k-2)!!}, & n = 2k \end{cases}.$$

Here, $n!! = n(n-2)(n-4)\ldots$

Proof. This is a well-known classical result. It could be derived from the frequency function for the square root $\sqrt{\sum_{i=1}^n x_i^2}$ given already in [4], Section 18.1. For completeness, we give a proof based on just elementary calculus.

Since the probability density function for each component x_j is $f(x_j) = \frac{1}{\sqrt{2\pi}} e^{-x_j^2/2}$ and the components are independent, we get by standard probability analysis:

$$E(\|x\|) = \frac{1}{(2\pi)^{n/2}} \int_{-\infty}^{\infty} \ ... \ \int_{-\infty}^{\infty} \|x\| e^{-\|x\|^2/2} dx_1 \ ... \ dx_n.$$

By changing to spherical coordinates we readily get:

$$\begin{aligned} E(\|x\|) &= \frac{1}{(2\pi)^{n/2}} \int_0^{\infty} \int_0^{2\pi} \int_0^{\pi} \ ... \ \int_0^{\pi} r^n e^{-r^2/2} \sin\phi_2 \sin^2\phi_3 \ ... \\ &\quad \sin^{n-2}\phi_{n-1} d\phi_{n-1} \ ... \ d\phi_2 d\phi_1 dr \\ &= \frac{2\pi}{(2\pi)^{n/2}} \int_0^{\infty} \int_0^{\pi} \ ... \ \int_0^{\pi} r^n e^{-r^2/2} \sin\phi_2 \sin^2\phi_3 \ ... \\ &\quad \sin^{n-2}\phi_{n-1} d\phi_{n-1} \ ... \ d\phi_2 dr. \end{aligned} \tag{3}$$

For evaluating the integrals in (3) we use some standard definite integrals, see for instance [10]:

$$\int_0^{\pi} \sin^n x \ dx = \begin{cases} \frac{2(n-1)!!}{n!!}, & n = 1, 3, 5, ... \\ \frac{\pi(n-1)!!}{n!!}, & n = 2, 4, 6, ... \end{cases} \tag{4}$$

and

$$\int_0^{\infty} x^n e^{-x^2/2} \ dx = \begin{cases} \frac{(2k-1)!!\sqrt{2\pi}}{2}, & n = 2k \\ k! 2^k, & n = 2k+1 \end{cases} . \tag{5}$$

By using (4) and (5) for the separate integrals in (3) the formulas for the expected norm arise. □

Theorem 4.1. *Let $x \sim N(0, I_n)$. Then for $\hat{x} = A^+ Ax$ we get the mean of the norm*

(i) $E(\|\hat{x}\|) = E(\|z_1\|)$, *where $z_1 \in \mathbb{R}^m$ and $z_1 = N(0, I_m)$.*

Further, for n and m even, the mean of the angle v between x and $\hat{x}$ is given by

(ii) $E(\cos v) = \frac{\pi(n-3)!!(m-1)!!}{2(n-2)!!(m-2)!!}$.

In the limit for large dimensions, we get with fixed $s = n/m$ the mean

(iii) $\lim_{m\to\infty} E(\cos v) = \sqrt{\frac{m}{n}} = \sqrt{\frac{1}{s}}$.

A similar formula to (ii) *is valid also for n and m being odd and for the mixed odd-even case.*

Proof. The SVD of $A \in \mathbb{R}^{m\times n}$, with full rank, can be written $A = U[\Sigma \ O]V^T$, where $\Sigma \in \mathbb{R}^{m\times m}$ is non-singular and O is the $m \times (n-m)$

zero matrix. The orthogonal projection P_{A^T} onto $Col(A^T) = Nul(A)^\perp$ can be expressed by

$$P_{A^T} = A^+A = V \begin{bmatrix} I_m & O \\ O & O \end{bmatrix} V^T.$$

Furthermore, we have $\hat{x} = P_{A^T}x = A^+Ax$. Consider the orthogonal transformation $z = V^Tx$. Since V is orthogonal, it follows that $z = N(0, I_n)$, see for instance [4], Section 24.4. Let z be splitted according to the SVD, i.e., $z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$. Then we may write $\hat{z} = V^T\hat{x} = \begin{bmatrix} z_1 \\ O \end{bmatrix}$. Since V is orthogonal, $E(\|\hat{z}\|) = E(\|\hat{x}\|)$, and apparently $E(\|\hat{z}\|) = E(\|z_1\|)$ so $E(\|\hat{x}\|) = E(\|z_1\|)$ and part (i) of the theorem is proved.

In order to prove part (ii) of the theorem we first calculate $E(\frac{1}{\|x\|})$ in a similar way that we calculated $E(\|x\|)$ in Lemma 4.1. We readily find that $E(\frac{1}{\|x\|})$ differs from $E(\|x\|)$ in (3) just by the factor r^{n-2} in place of r^n in the integral and we conclude that $E(\frac{1}{\|x\|}) = E(\|x\|)/(n-1)$. Note that $\cos v$ is independent of the norm $\|x\|$. Using part (i) we now get

$$E(\cos v) = E(\frac{\|\hat{x}\|}{\|x\|}) = E(\frac{1}{\|x\|})E(\|\hat{x}\|) = E(\|z_1\|)E(\|x\|)/(n-1),$$

where $x = N(0, I_n)$ and $z_1 = N(0, I_m)$.

From Lemma 4.1 we then get, for n and m even, the desired result (ii):

$$E(\cos v) = \frac{\pi(n-3)!!(m-1)!!}{2(n-2)!!(m-2)!!}.$$

For part (iii), we use asymptotic behaviors (when $n \to \infty$) of the gamma function

$$\Gamma(\frac{n-1}{2}) = \Gamma(\frac{n-3}{2}+1) \sim \sqrt{2\pi}(\frac{n-2}{2})^{\frac{n-2}{2}}e^{-(\frac{n-2}{2})} \tag{6}$$

and for the factorial

$$(n-2)! \sim \sqrt{2\pi}(n-2)(\frac{n-2}{e})^{n-2} \tag{7}$$

see for instance [10].

We also use the well-known relation between !! and the gamma function, see for instance [1]:

$$(n-3)!! = \Gamma(\frac{n-1}{2})2^{\frac{n-2}{2}}/\sqrt{\pi}, \tag{8}$$

By the equations (6), (7) and (8) we now get

$$\frac{(n-3)!!}{(n-2)!!} = \frac{((n-3)!!)^2}{(n-2)!} \sim \frac{2^{n-2}2\pi(\frac{n-2}{2})^{n-2}e^{-2(\frac{n-2}{2})}e^{n-2}}{\pi\sqrt{2\pi(n-2)}(n-2)^{n-2}}$$

$$= \frac{2^{n-1}(\frac{n-2}{2})^{n-2}}{\sqrt{2\pi(n-2)}(n-2)^{n-2}} = \frac{\sqrt{2}}{\sqrt{\pi(n-2)}}. \tag{9}$$

From (ii) we have

$$E(\cos v) = \frac{\pi(n-3)!!(m-1)(m-3)!!}{2(n-2)!!(m-2)!!},$$

and by using (9) for n and m, respectively, the desired result (iii) for fixed $s = n/m$ with $m \to \infty$ is derived as follows:

$$E(\cos v) \sim \frac{m-1}{\sqrt{(n-2)(m-2)}} \sim \sqrt{\frac{m}{n}} = \sqrt{\frac{1}{s}}.$$

□

4.2. Examples. For instance, if we have twice as many columns than (independent) rows, then $s = 2$ and $\cos v$ is $\frac{1}{\sqrt{2}} \approx 0.7$. This means that we expect the relative gain, compared to steepest descent, to be 70% with 50% of the effort if we take a reasonably large step in the direction of $\hat{x}$. This result is proven only for a normally distributed x but it seems to be valid also if the components of x are equally distributed random numbers in the interval $(0,\ 1)$, see Figure 4 below.

It follows that the expected relative gain per parameter is $\sqrt{\frac{m}{n}}$ (when A has full rank). The gain for more columns is proportional to the number of columns so the total gain for a full rank investigation with supersaturated design on a linear system is $\sqrt{\frac{m}{n}} \cdot n = \sqrt{m \cdot n}$. In the case of not full rank, m should be replaced by $r = rank(A)$ in these formulas. This says that including some extra parameters is always a gain when doing an investigation.

In Figure 2 we present the expected relative gain as a function of $1/s$, where $1/s = m/n$ in the full rank case. We see that the less experiment we do, the bigger is the relative gain per experiment. We can see, for instance, that with a saturation of 4 one gets 50% of the relative gain instead of expected 25%.

Figure 3 shows the expected value of $\cos v = \|\hat{x}\|/\|x\|$ for different saturations and different ranks of the design matrix. We use the results from Theorem 4.1 for drawing this graph. It is observed that the rank needs to be at least 5 to get close to the full effect of the method.

Figure 4 shows the spread in lean optimization simulations compared to expected value. Ten experiments (rows) are used for 100000 parameters (columns), i.e., the saturation is 10000. 1000 simulations were done with equally distributed coefficients in the interval $(-1,\ 1)$ and we see that the lowest relative gain is about 0.004. In average we gain about 100 times the value $\frac{m}{n} = \frac{1}{s}$ (in the full rank case). The result, proved in Theorem 4.1 for normally distributed coefficients, says that the expected relative gain should

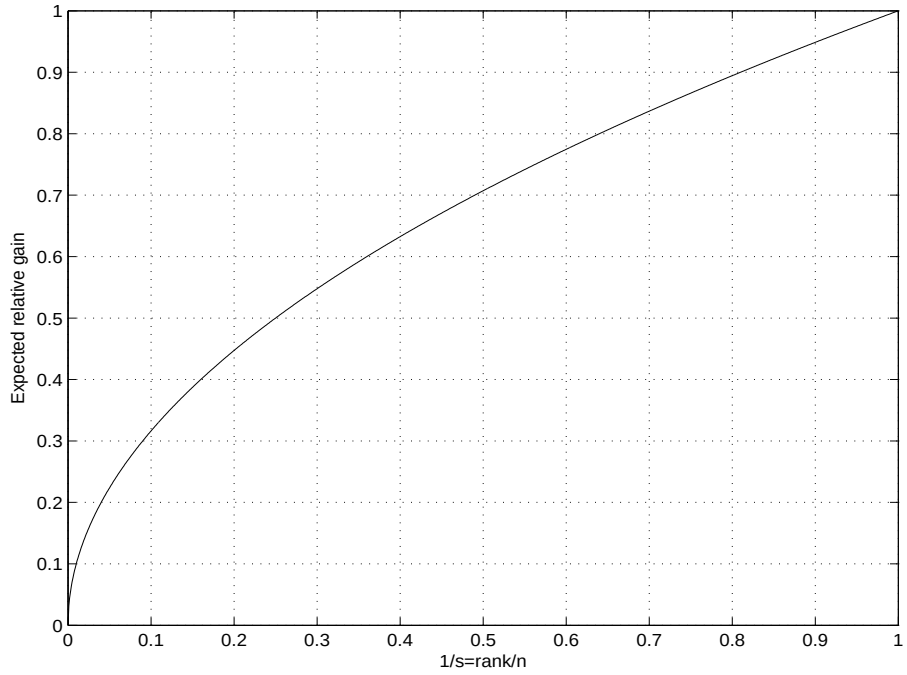

FIGURE 2. Expected relative gain per parameter in Supersaturated Design as a function of $1/s$, where $s = n/m$ is the saturation and the coefficients are normally distributed

be $\sqrt{\frac{1}{s}} = \sqrt{\frac{m}{n}} = 0.01$ and we conclude from this experiment that the same expected value seems to be valid also if the components of x are equally distributed random values in the interval $(-1,\ 1)$.

Since the null space of C has dimension at least $n - m > 0$, there always is a possibility that $\|Cx\| = 0$. This happens if $x \in Nul(A)$, and in this case one gets no information from the experiments. However, the expected value of the information is, relative to the number of experiments, higher than for a saturated experimental series.

For completeness we include, in Appendix 2, a MATLAB-function for testing the main result of Theorem 4.1, namely part (iii). This program was run for producing Figure 4.

5. Conclusions and discussion

The conclusions of this paper are as follows:

- Our Volvo data do not follow the Pareto principle.
- In SSD the rows should be linearly independent, not the columns.

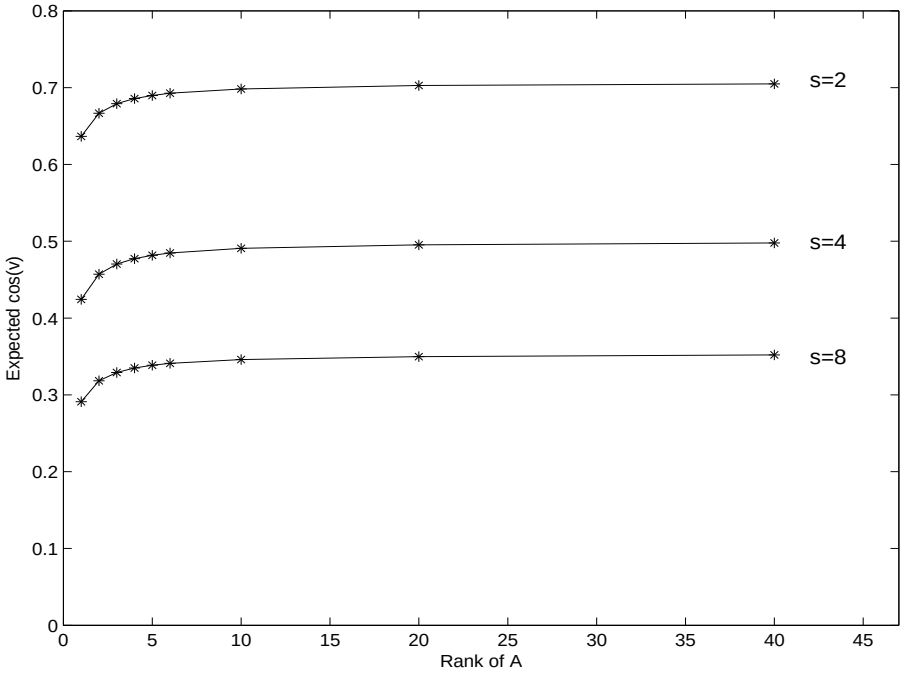

FIGURE 3. The expected value of $\cos v$ as a function of the rank of A for normally distributed coefficients and different values of the saturation s

- The expected relative gain per parameter with SSD is $\frac{1}{\sqrt{s}}$, where $s = \frac{n}{r}$, n is the number of columns (parameters) and r is the rank of the design matrix. For the full rank case, this gain is $\sqrt{\frac{m}{n}}$, where m is the number of rows (experiments).
- The total gain in lean optimization is expected to be $\sqrt{\frac{m}{n}} \cdot n = \sqrt{m \cdot n}$, in the full rank case.
- SSD is good only for reasonably many experiments, say $m > 5$.

One can understand our result as that the first observation gives the most information and then the information per observation decreases.

It is some kind of symmetry that nonsaturated design matrices A should have independent columns and supersaturated design matrices A should have independent rows.

Based on this article we conclude that the usual trend in DoE approach, to have more rows than columns, is not necessarily most efficient.

Our conclusion is that an experimental series should include as many parameters as are possible to investigate (change) and make a lean design

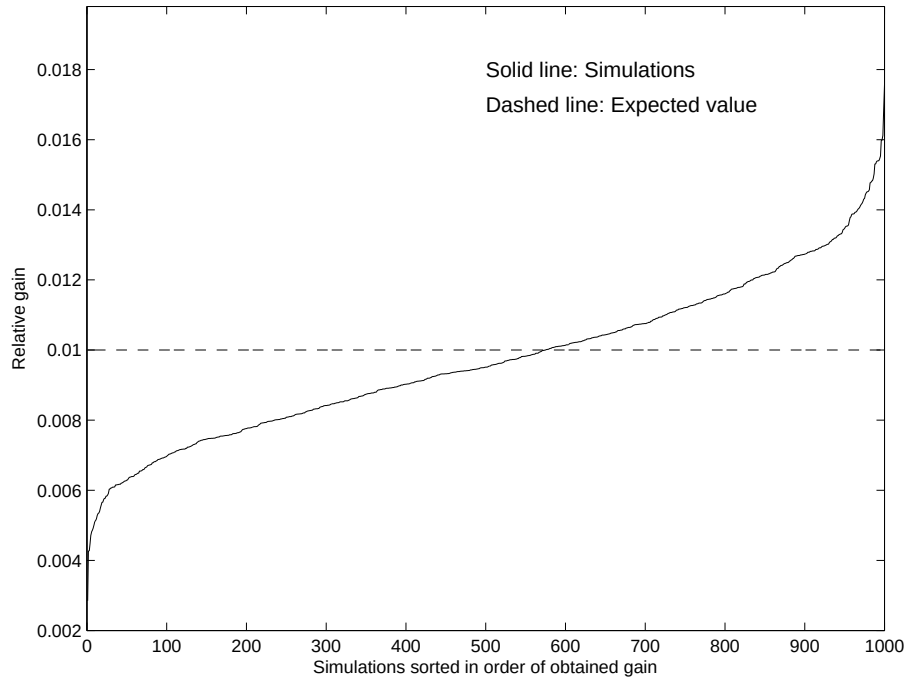

FIGURE 4. The relative gain per parameter for 100000 equally distributed cofficients in (-1,1) and ten experiments i.e., $s = 10000$. 1000 simuations are sorted in order of obtained gain.

for the experimental series. The direction of improvement is not the steepest descent but is still much better than m/n.

To further improve the product in a new experimental series, which can be afforded in the lean case, the lean design should be rotated which means the columns should be scrambled. Probably there is an optimal method to rearrange the columns for many experimental series, but this is outside the scope of this article.

In this study, perturbations have not been included. Probably such a study would result in that the rows in a supersaturated A should not only be independent but orthogonal to minimize the influence of the perturbations on b.

Appendix 1: Algorithm for AG-design

```
function L=AG_design(m,optional_n)
%Ahlinder-Gustafsson design
%m is the number of experiments in lean design, m > 4 to get it supersaturated.
```

```
%optional_n is optional number of variables
%The rank of the design matrix is m-1.

%Full factor design
M=[-1,1]';
for i=2:m % Loop for all variables except the first
M=[M;M]; %Stacking two matrices M
M=[M,[-ones(size(M,1)/2,1);ones(size(M,1)/2,1)]];
%Add a column with upper half being -1's and lower half being 1's
end;

%Full factor design to supersaturated design
v=2^m; %The number of rows is m
L=M(v/2+1:v,:); %Last half to avoid total correlation
L=L'; %Transpose to get it supersaturated
s=sum(L); %The sum of variables should be -1,0 or 1
L=L(:,or(s==0,s==1)); %To get balanced designs
if nargin==2 %If the number of arguments is =2
L=L(:,1:optional_n); %Cut from end if optional_n is given.
end
```

Appendix 2: The function leantest

```
  function leantest
%Test of Ahlinder-Gustafssons theorem
%Solve a system of n varables with m rows and do it p times
m=10; %Number of rows each time
n=10000; %Number of variables each time
p=1000; %Number of sets run
c=zeros(p,1); %Correlation each time
A=randn(m,n); %The supersaturated design
C=pinv(A)*A; %The cap matrix, xhat=Cx
for i=1:p
x=randn(n,1); %The correct x
xhat=C*x; %The estimated x
c(i)=norm(x'*xhat)/(norm(x)*norm(xhat)); %correlation
end
plot(1:p,sort(c))
e=sqrt(m/n); %expected correlation
hold on
plot([0,p],[e,e],':')
plot(p/2,e,'*')
```

Acknowledgements

We want to thank Volvo Technology cooperation for financing this reserch. We also want to thank former Volvo Lastvagnar for financing the basic training. Further, we would like to express our gratitude to professor Åke Björck for valuable discussions regarding the proof of Theorem 4.1. Finally, we also want to thank an anonymous referee for comments and suggestions for the improvement of the paper.

References

[1] G. Arfken, *Mathematical Methods for Physicists*, 3rd edition, Academic Press, San Diego, 1985, p. 548.
[2] G. E. P. Box, W. G. Hunter, and J. S. Hunter, *Statistics for Experimenters*, Wiley Interscience, New York, 1978.
[3] Ching-Shui Cheng, *$E(s^2)$-optimal supersaturated designs*, Statist. Sinica **7** (1997), 929–939.
[4] H. Cramér, *Mathematical Methods of Statistics*, Almqvist and Wiksell, Stockholm, 1945.
[5] J. W. Demmel, *Applied Numerical Linear Algebra*, SIAM, Philadelphia, 1997.
[6] R. A. Fisher, *The Design of Experiments*, Macmillan Pub Co, New York, 1971.
[7] O. Goury, *On efficient black box optimizition of systems defined by 100 or more parameters*, Master's thesis, Department of Mathematical Sciences, Chalmers University of Technology, Gothenburg, Sweden, 2009.
[8] D. J. K. Lin, *Generating systematic supersaturated designs*, Technometrics **37** (1995), 213–225.
[9] *Pareto principle*, http://en.wikipedia.org/wiki/Pareto_principle, January 2012.
[10] L. Råde and B. Westergren, *Mathematics Handbook for Science and Engineering*, Springer-Verlag, Berlin; Studentlitteratur, Lund, 2004.
[11] I. Siomina and S. Ahlinder, *Lean optimization using supersaturated experimental design*, Appl. Numer. Math. **58** (2008), 1–15.
[12] *The hat matrix*, http://en.wikipedia.org/wiki/Hat_matrix, January 2012.

Volvo GTT Advanced Technology and Research, Dept BF40510 CTP9A, SE-405 08 Gothenburg, Sweden
E-mail address: sven.ahlinder@volvo.com

Department of Mathematical Sciences, SE-412 96 Gothenburg, Sweden
E-mail address: ivar@chalmers.se