



Equitable (d, m) -edge designs

Jean-Marc Fédou, Maria João Rendas

► To cite this version:

| Jean-Marc Fédou, Maria João Rendas. Equitable (d, m) -edge designs. 2013. hal-01318800

HAL Id: hal-01318800

<https://hal.science/hal-01318800>

Preprint submitted on 20 May 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Equitable (d, m) -edge designs

Jean-Marc Fédou*, Maria-João Rendas**

Laboratoire I3S - UMR7271 - UNS CNRS 2000, route des Lucioles 06900 Sophia Antipolis - France

Abstract

The paper addresses design of experiments for classifying the input factors of a multi-variate function into negligible, linear and other (non-linear/interaction) factors. We give constructive procedures for completing the definition of the clustered designs proposed in [10], that become defined for arbitrary number of input factors and desired clusters' multiplicity. Our work is based on a representation of subgraphs of the hyper-cube by polynomials that allows the formal verification of the designs' properties. Ability to generate these designs in a systematic manner opens new perspectives for the characterisation of the behaviour of the function's derivatives over the input space that may offer increased discrimination.

Keywords: Sensitivity analysis, clustered designs, one at a time designs.

1. Introduction

1.1. Sensitivity analysis

In sensitivity analysis, one wishes to characterise the dependency of an unknown function $f : \mathcal{A} \subset \mathbb{R}^d \rightarrow \mathbb{R}$ on each of its d input factors. In general, we know nothing about the function $f(\cdot)$, but can evaluate it at chosen locations $\xi \in \mathbb{R}^d$. Interest is on partitioning the factors of $f(\cdot)$ into those that have no impact on the function value (class \mathcal{C}_0), that have a linear effect (class \mathcal{C}_1) or that are non-linear or have interactions with other input factors (class \mathcal{C}_2). Often, fast screening is done in the context of factor

*Principal corresponding author

**Corresponding author

Email addresses: fedou@unice.fr (Jean-Marc Fédou), rendas@i3s.unice.fr (Maria-João Rendas)

fixing (as noted in [13]), with the goal of restricting subsequent analysis of $f(\cdot)$ to the smaller set \mathcal{C}_2 . This is the context we address.

Several methods have been proposed for sensitivity analysis, ranging from local to global methods, in particular, variance based methods such as the use of Monte-Carlo methods for the computation of Sobol indices ([15]), the Fourier Amplitude Sensitivity Test (FAST) method ([6], [7]), or the Morris elementary effect method ([10]). Morris method for preliminary sensitivity analysis is one of the most commonly used, due to its robustness and computational efficiency. The method has not only been applied to a variety of different fields (see [13] for a review), but has also received the attention of several researchers who proposed modifications and improvements: [5]–[2] propose an extension enabling study of two-factor interaction terms, [3] chooses the design used to evaluate the elementary effects amongst a large number of random trajectories, such that a dispersion index is optimised, [12] replaces designs aligned with the input space directions by randomly oriented simplexes, and [1] proposes a sequential version of Morris test, so that computational effort is concentrated in class \mathcal{C}_2 factors. We focus on Morris’ original method, that we outline below. Our contribution concerns the designs used in Morris method, and can be combined with most modifications of the original method published in the literature.

1.2. Morris preliminary sensitivity analysis designs

Morris method implements statistical tests over a set of elementary differences along each principal direction i , $d_i(\xi)$, computed at a set of points $\{\xi_n\}_{n=1}^r$ of the input domain:

$$d_i(\xi) = \frac{1}{\Delta} [f(\xi + \Delta e_i) - f(\xi)], \quad \xi \in \mathcal{A}, \quad i = 1, \dots, d. \quad (1)$$

Above, e_i is the vector with components $e_{ij} = \delta_{ij}$, $j = 1, \dots, d$. Let (μ_i, σ_i^2) be empirical estimates of the mean \bar{d}_i and variance s_i^2 of d_i :

$$\mu_i = \frac{1}{r} \sum_{n=1}^r d_i(\xi_n^i) \simeq \mathbb{E}_{\nu_i} [d_i] = \bar{d}_i, \quad i = 1, \dots, d \quad (2)$$

$$\sigma_i^2 = \frac{1}{r-1} \sum_{n=1}^r (d_i(\xi_n^i) - \mu_i)^2 \simeq \text{Var}_{\nu_i} [d_i] = s_i^2. \quad (3)$$

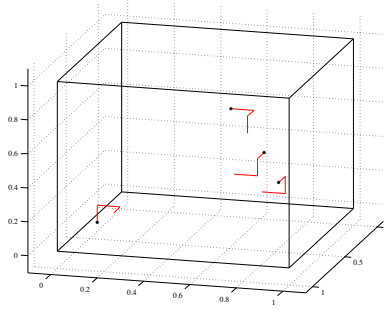


Figure 1: Illustration of Morris elementary effects method ($d = 3, r = 4$).

Input factors are classified as (i) negligible, (ii) linear, or (iii) non-linear/interaction if (i) their mean and variance are both close to zero, (ii) the mean is non-zero, but variance is small, or (iii) variance is large. A revised version of Morris method ([3]) uses instead μ_i^* , the sample average of $|d_i(\xi)|$, improving the robustness for derivatives of alternating sign.

If points $\{\xi_n^i\}_{n=1, i=1}^{r, d}$ are chosen completely at random, the sensitivity analysis of a function of d variables requires a total of $2dr$ evaluations of $f(\cdot)$. The basic Morris scheme is a One-At-a-Time (OAT) method that increases efficiency with respect to random sampling by using most evaluations of $f(\cdot)$ twice. It relies on empirical moment estimates using r samples of $\{d_i(\cdot)\}_{i=1}^d$ computed along r randomly oriented paths T_{d+1} along which each one of the d coordinates is changed at a time, see Figure 1. The total number of evaluations of $f(\cdot)$ is $r(d+1)$, which for large values of r and d may still be prohibitive. Morris clustered designs, see Section 5 in [10], improve on the efficiency of these OAT designs by using each value of f in the computation of more than two elementary differences. The simple paths T_{d+1} are replaced by denser graphs that enable determination of $m \geq 1$ elementary differences along each direction.

For large values of d , Morris relies on a factorisation of the input space. Let Q_d be the d -dimensional unit hypercube, and factor $d = tq$ such that $Q_d = Q_q^t$. Let $Y = \{\xi_1, \dots, \xi_{|Y|}\}$ be a design in $Q_q(\xi)$ that enables the determination of m elementary

effects along each direction. The full design $\Xi = \bigcup_{j=1}^t Y_{(j)}$, where $Y_{(j)}$ is a replication of Y along coordinates $X_{(j-1)q+1}, \dots, X_{jq}$, computes at least m elementary effects along each coordinate.

Although this idea is interesting, Morris' presentation is affected by a number of drawbacks. In [10] the smaller designs $Y \subset Q_q$ gather all $s \in Q_q$ with ℓ bits equal to one for all $\ell \in \mathcal{I} \subset \{0, \dots, q\}$. Design multiplicity m indirectly follows from choice of \mathcal{I} , but no guidelines on how this list should be chosen are provided, and actually, since not all integers can be decomposed as the sum of a set of powers of two, not all multiplicities $m \leq 2^{d-1}$ can be obtained. Note also that d must not be prime and

$$d = tq \geq 2q_{\min}(m) = 2\lceil \log_2(m) \rceil + 2, \quad q \geq q_{\min}(m) = \lceil \log_2(m) \rceil + 1. \quad (4)$$

Verification of the properties of Morris' clustered designs is cumbersome and their optimality, as it is recognised by the author, is not guaranteed. In fact, since Morris designs are not necessarily connected – they will be if $q \in \mathcal{I}$ – they are not natural candidates for optimality.

1.3. Contributions

The main result of the paper is the explicit presentation of a family of subgraphs of Q_d that enable the computation of a pre-specified number m of elementary effects for all $1 \leq m \leq 2^{d-1}$.

Definition 1. A subgraph S of Q_d is a (d, m) -edge equitable design if and only if the number of edges of S along each direction is exactly m .

Figure 2 illustrates this definition. For each graph, edge colour indicates the direction of Q_d along which the edge is aligned (we will use this colour code for arbitrary values of d). The number of edges of each colour is thus exactly equal to m for (d, m) -edge equitable graphs. The graph on the left is $(3, 2)$ -edge equitable, while the other two graphs are not edge equitable.

Although a vast literature characterising interesting families of subgraphs of the hypercube, such as median and meshed subgraphs, as well as on graph colouring problems, exists in discrete mathematics, see e.g. [8], we could find no reference addressing this class of subgraphs, and their determination seems to be largely an open problem.

We present recursive procedures (Algorithms 1, 2 and 3) that generate (d, m) -edge equitable designs, overcoming most of the drawbacks of Morris' construction: (i) they are guided by the values of m and d , (ii) handle generic values of (d, m) , and (iii) provably lead to equitable designs.

We claim an additional contribution, that consists in the exploitation of a convenient polynomial representation of subgraphs of Q_d . A related map between polynomials and subgraphs of Q_d , the log map, has been used in [11] to study the class of polynomial models identifiable by a design, using computational commutative algebra. We believe that the polynomial representation of subgraphs of Q_d and, more importantly, the exploitation of a suitably defined scalar product over polynomials for formal verification of several graph properties, without having to resort to intricate combinatorial arguments, is novel. In particular, we are able to provide algebraic demonstrations for equitability (Theorem 2), and derive explicit formulas for the size of our designs (Theorems 3, 4 and 5).

As the paper shows, improved efficiency in the computation of the elementary effects by using clustered designs does not translate into better performance on the classification of input factors in Morris original method. Definition of tests adapted to the structured sampling implemented by clustered designs will be addressed in a forthcoming paper.

2. Polynomial representation of subgraphs of Q_d

We concentrate on subgraphs of the unit hypercube $Q_d = \{0, 1\}^d$, i.e., the graph whose vertices are the points having coordinates 0 or 1 in \mathbb{R}^d , two points being joined by an edge if and only if they differ in exactly one coordinate. Given an ordering of the directions of Q_d , there is a bijection between its vertices and the binary words of length d :

$$\begin{aligned} Q_d &\rightarrow \{0, 1\}^d \\ s &\mapsto \{s_i\}_{i=1}^d, s_i \in \{0, 1\}. \end{aligned}$$

We define a d -edge-coloring of Q_d by stating that an edge joining two points s and s' has color i when $s_i \neq s'_i$ and $s_j = s'_j, j \neq i$.

We associate to each $s \in Q_d$ a monomial \mathcal{P}_s in the ring $\mathbb{R}[X_1, \dots, X_d]$ of the polynomials over the variables X_1, \dots, X_d :

$$s = \{s_1, \dots, s_d\} \longrightarrow \mathcal{P}_s(X_1, \dots, X_d) = X_1^{s_1} \dots X_d^{s_d} .$$

The subgraph induced by a set $S \subset Q_d$ will be represented by the polynomial $\mathcal{P}_S = \sum_{s \in S} \mathcal{P}_s$. The empty set is represented by the zero polynomial. The set of the polynomials representing simple subgraphs of Q_d will be denoted by K_d , and corresponds to the polynomials of degree at most 1 in each variable having coefficients in $\{0, 1\}$.

2.1. Scalar product in K_d

The set K_d can be embedded in the algebra $\mathbb{R}[X_1, \dots, X_d]/\{X_i^2 \equiv 1, i = 1 \dots d\}$ of the polynomials according to the equivalence relation induced by the equalities $X_i^2 \equiv 1, i = 1 \dots d$. This algebra is a vector space for which the set of monomials can be taken as a natural basis. By defining a scalar product such that this basis is orthogonal, we endow K_d with a structure that has several interesting properties in term of the underlying subgraphs of Q_d .

Definition 2. We define the scalar product between monomials $\mathcal{P}_s, \mathcal{P}_{s'} \in K_d$ as

$$\langle \mathcal{P}_s, \mathcal{P}_{s'} \rangle = 1_{s=s'} ,$$

and extend it naturally to the entire K_d by bilinearity

$$\langle \mathcal{P}_S, \mathcal{P}_{S'} \rangle = \sum_{s \in S, s' \in S'} \langle \mathcal{P}_s, \mathcal{P}_{s'} \rangle, \quad \mathcal{P}_S, \mathcal{P}_{S'} \in K_d .$$

Lemma 1. *The scalar product of two subgraphs of Q_d , S_1 and S_2 , is equal to the size of their intersection: $\langle \mathcal{P}_{S_1}, \mathcal{P}_{S_2} \rangle = |S_1 \cap S_2|$. In particular, $\langle \mathcal{P}_S, \mathcal{P}_S \rangle = \|\mathcal{P}_S\|^2 = |S|$.*

Lemma 2. *Let $s \in Q_d$ and $S \subset Q_d$. The subgraph S' defined by $\mathcal{P}_{S'} = \mathcal{P}_s \mathcal{P}_S$ is the reflection of S along the directions present in s . In particular, $X_i \mathcal{P}_S$ corresponds to the mirror of S along direction i .*

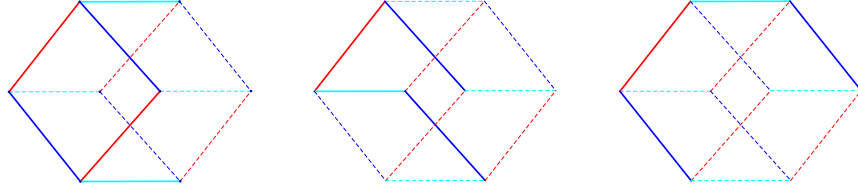


Figure 2: Edge equitable (left) and non edge equitable (centre) graphs. The colour of each edge indicates the direction along which it is aligned. The graph on the right is the reflection of the graph on the left along X_1 .

Figure 2 illustrates Lemma 2, showing $\mathcal{P}_S = 1 + X_1 + X_2 + X_1X_3 + X_2X_3$ and $X_1\mathcal{P}_S$. Multiplication by X_1 resulted in a reflection of S along the red (X_1) direction.

Lemma 3. *For all $s \in Q_d$, $S, S' \subset Q_d$ $\langle \mathcal{P}_s \mathcal{P}_S, \mathcal{P}_s \mathcal{P}_{S'} \rangle = \langle \mathcal{P}_S, \mathcal{P}_{S'} \rangle$.*

Using Lemmas 2 and 1 the following is immediate.

Lemma 4. *The number m_i of edges of $S \subset Q_d$ having color i satisfies*

$$\langle \mathcal{P}_S, X_i \mathcal{P}_S \rangle = 2m_i, \quad i \in \{1, \dots, d\} . \quad (5)$$

2.2. Problem (re)formulation

Denote by \mathcal{E}_m^d the set of (d, m) -edge equitable polynomials. Using Lemma 4,

Lemma 5. $\mathcal{P}_S \in \mathcal{E}_m^d$ if and only if

$$\langle \mathcal{P}_S, X_i \mathcal{P}_S \rangle = 2m, \quad i \in \{1, \dots, d\} . \quad (6)$$

Lemma 6. \mathcal{E}_m^d is closed under multiplication by monomials:

$$\forall s \in Q_d, \quad \mathcal{P}_S \in \mathcal{E}_m^d \Rightarrow \mathcal{P}_{S'} = \mathcal{P}_s \mathcal{P}_S \in \mathcal{E}_m^d .$$

and under permutations of the coordinates of Q_d .

Theorem 1. *Let $S \subset Q_d$, and \bar{S} denote the complement of S in Q_d .*

$$\mathcal{P}_S \in \mathcal{E}_m^d \Rightarrow \mathcal{P}_{\bar{S}} \in \mathcal{E}_{2^{d-1}+m-|S|}^d ,$$

i.e., the complement of a (d, m) -edge equitable subgraph is an (d, m') -edge equitable graph, with $m' = 2^{d-1} + m - |S|$.

Proof.

Let $\mathcal{P}_S \in \mathcal{E}_m^d$ and compute $\langle \mathcal{P}_{\overline{S}}, X_i \mathcal{P}_{\overline{S}} \rangle$

$$\begin{aligned} \langle \mathcal{P}_{\overline{S}}, X_i \mathcal{P}_{\overline{S}} \rangle &= \langle \mathcal{P}_{Q_d} - \mathcal{P}_S, X_i (\mathcal{P}_{Q_d} - \mathcal{P}_S) \rangle \\ &= \langle \mathcal{P}_{Q_d}, X_i \mathcal{P}_{Q_d} \rangle + \langle \mathcal{P}_S, X_i \mathcal{P}_S \rangle - 2 \langle \mathcal{P}_{Q_d}, X_i \mathcal{P}_S \rangle = 2^d + 2m - 2|S| = 2m', \end{aligned}$$

which is independent of i , completing the proof.

Theorem 1 is a first illustration of the power of the polynomial representation for establishing the properties of subgraphs of the hypercube.

3. Generation of (d, m) -equitable subgraphs of Q_d

3.1. Recursive graph composition

For every natural number d and every integer $1 \leq m \leq 2^{d-1}$ the algorithm below produces a polynomial $G_m^d \in \mathcal{E}_m^d$.

Algorithm 1. Recursive definition of G_m^d .

Initialization ($m = 1$):

$$G_1^d = 1 + \sum_{i=1}^d X_i. \quad (7)$$

Recursion

$$\bullet \text{ For } m \text{ even, } \quad G_m^d = (1 + X_1 X_d) G_{m/2}^{d-1}. \quad (8)$$

$$\bullet \text{ For } m \text{ odd, } \quad G_m^d = G_{(m-1)/2}^{d-1} + X_1 X_d G_{(m+1)/2}^{d-1}. \quad (9)$$

Figure 3 illustrates the graph compositions of eq. (8) (m even) and eq. (9) (m odd), respectively. Note that in the graphs displayed on the right the m edges linking the two graphs on the left are along the new dimension X_d (green colour). The solutions are the composition of graphs with smaller values of d and m , along a binary tree whose leaves all have $m = 1$.

Theorem 2. For all $d \geq 1$, and all $1 \leq m \leq 2^{d-1}$, the graphs G_m^d defined by (7) – (9) are (d, m) -edge equitable.

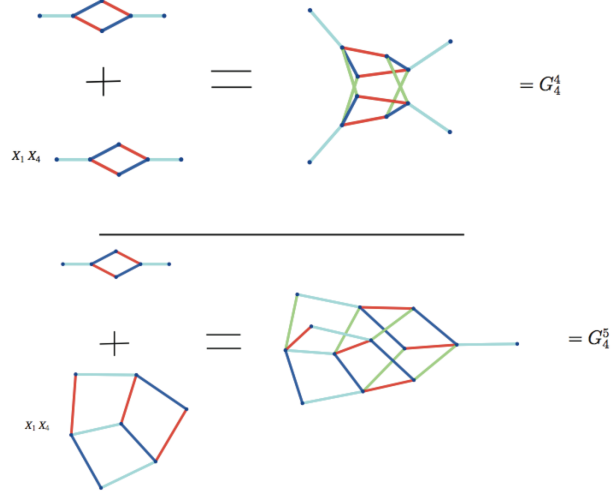


Figure 3: Construction of G_4^4 (top) and of G_5^4 (bottom).

The proof of Theorem 2 is given in Appendix A, and is based on the Lemmas of Section 2. Theorem 2 overcomes major limitations of Morris presentation, by defining a solution (i) for all pairs (d, m) which is (ii) provably equitable.

Theorem 3. *For the graphs G_m^d defined by (7) – (9)*

$$|G_m^d| = m(d - \kappa) + 2^{\kappa+1} - m, \quad \text{where } \kappa = \lfloor \log_2(m) \rfloor. \quad (10)$$

Demonstration of Theorem 3 is trivial by verifying that (10) is satisfied if we initialise with $|G_1^d| = d + 1$ the recurrence equations implied by (8) and (9)

$$\begin{aligned} |G_m^d| &= 2 |G_{m/2}^{d-1}|, & \text{for } d \text{ even} \\ |G_m^d| &= |G_{m-1/2}^{d-1}| + |G_{m+1/2}^{d-1}|, & \text{for } d \text{ odd}. \end{aligned}$$

3.2. Improving efficiency by an alternative initialisation

Since $|G_m^d|$ is recursively defined, decreasing size for small values of m will propagate to larger values of m . We present now an alternative family of (d, m) -edge

equitable graphs H_m^d , $m \geq 2$, obtained by starting the recursion (8) – (9) at $m = 4$. The minimality of these graphs can be proved for $m = 2$, and has been checked numerically for $m = 3, d \leq 5$.

Algorithm 2. *Recursive definition of H_m^d .*

Initialization ($m = 2, 3$)

- For $m = 2$ we distinguish the cases of even and odd d :

$$\text{when } d \geq 2 \text{ is even} \quad H_2^d = H_2^{d-2} + (X_{d-1} + X_d + X_{d-1}X_d), \quad (11)$$

$$\text{when } d \geq 3 \text{ is odd,} \quad H_2^d = H_2^{d-1} + X_1X_d + X_{d-1}X_d. \quad (12)$$

- For $m = 3$, $H_3^d = 1 + X_1X_d + \sum_{k=1}^d X_k + \sum_{j=1}^{d-1} X_jX_{j+1}$.

Recursion: Apply eqs. (8) – (9).

The size of these graphs satisfies the recursive equations

$$|H_2^d| = \begin{cases} |H_2^{d-2}| + 3(d-2), & \text{if } d \text{ is even} \\ |H_2^{d-3}| + 5, & \text{if } d \text{ is odd} \end{cases} \quad (13)$$

$$|H_3^d| = 1 + 2d$$

By writing $m \geq 3$ as $m = 2p_2 + 3p_3$, $p_2, p_3 \in \mathbb{N}^0$, where p_2 and p_3 are the number of leaves labeled 2 and 3, respectively, in the recursive decomposition of m used in our algorithm, the following Lemma can be demonstrated:

Lemma 7. *Let $k = \lfloor \log_2(m) \rfloor$, and write $m = 2^\kappa + 2^{\kappa-1} + i$, $i \in [-2^{\kappa-1}, 2^{\kappa-1}]$. Then the number of subgraphs H_2^* and H_3^* in the recursive composition (8)-(9) are, respectively,*

$$p_2 = \begin{cases} 2i, & \text{if } i \geq 0 \\ -i, & \text{if } i < 0 \end{cases}, \quad p_3 = 2^{\kappa-1} - |i|.$$

If $i < 0$ all the subgraphs are in dimension $d - \kappa + 1$, otherwise the subgraphs with $m = 2$ are in dimension $d - \kappa$.

Theorem 4. Let κ and i be defined as in Lemma 7. The size of H_m^d is

$$\begin{aligned} |H_m^d| &= c(m) + \alpha(m)d, \\ \alpha(m) &= \begin{cases} i + 2^\kappa, & i \geq 0 \\ i/2 + 2^\kappa, & i \leq 0 \end{cases} = \begin{cases} m - 2^{\kappa-1}, & i \geq 0 \\ \frac{1}{2}(m + 2^{\kappa-1}), & i < 0 \end{cases}, \end{aligned} \quad (14)$$

where the term independent of d is

$$c(m) = \begin{cases} -m \left(\frac{1}{2}((-1)^{d-\kappa} + 1) + \kappa \right) + 2^{\kappa-2} (3(-1)^{d-\kappa} + 2\kappa + 9), & i \geq 0 \\ -\frac{1}{2}m \left(\frac{1}{2}((-1)^{d-\kappa} - 1) + \kappa \right) - 2^{\kappa-3} (-3(-1)^{d-\kappa} + 2\kappa - 9), & i < 0 \end{cases}$$

Proof is simple by verifying the recursive equations (13).

For large values of κ , it can be shown that

$$|H_m^d| \simeq \begin{cases} (d - \kappa)(m - 2^{\kappa-1}), & i \geq 0 \\ \frac{1}{2}(d - \kappa)(m + 2^{\kappa-1}), & i < 0 \end{cases}$$

When $m = 2^\kappa + 2^{\kappa-1}$ a simpler expression can be found:

$$|H_{2^\kappa + 2^{\kappa-1}}^d| = 2^\kappa \left((d - \kappa) + \frac{3}{2} \right). \quad (15)$$

Theorem 4 completes the characterisation of the family of solutions $\overline{H}_{m'}^d = \overline{H}_m^d$, in terms of their size, which is $|\overline{H}_{m'}^d| = 2^d - |H_m^d|$, and of the value of m' .

3.3. Further improving economy by factoring the designs

Consider the case $m = 2^\kappa + 2^{\kappa-1}$ when the simpler expression in (15) holds. We can check that

$$|H_m^{2d}| = 2|H_m^d| + 2^\kappa \left(\kappa - \frac{3}{2} \right) > 2|H_m^d|,$$

i.e., the size of our designs grows supra-linearly in d . It can be checked that this is true for generic values of m . We improve the family of designs presented in the previous subsection by combining the factorisation approach used by Morris clustered designs with the generic solution H_m^d presented in the previous section.

Remember the definition of $q_{\min}(m) = \lceil \log_2(m) \rceil + 1$, see equation (4), as the dimension of the smallest hypercube that can contain m edges along each direction, and for a given pair (d, m) write d as $d = c \cdot q_{\min}(m) + t'$, such that

$$d = (c - 1)q_{\min}(m) + t, \quad t \in \{q_{\min}(m), \dots, 2q_{\min}(m) - 1\}. \quad (16)$$

It is easy to check that $q_{\min}(m) = \kappa + 1$, where κ is the parameter in Lemma 7. For $d \geq 2q_{\min}(m)$ designs more efficient than those presented in section 3.2 can be obtained by placing $c - 1$ copies of $H_m^{q_{\min}(m)}$ in disjoint $q_{\min}(m)$ -dimensional subspaces of Q_d , and adding a H_m^t design covering the remaining directions. In the following we will often omit indication of the dependency on m , using the simpler notation q_{\min} .

Algorithm 3. Definition of M_m^d .

$$M_m^d = 1 + \sum_{j=1}^{c-1} [\text{Shift}_{jq_{\min}}(H_m^{q_{\min}}) - 1] + \text{Shift}_{(c-1)q_{\min}} H_m^t, \quad (17)$$

where $\text{Shift}_k(\mathcal{P})$ operates over the coordinates of the polynomial \mathcal{P} :

$$\text{Shift}_k(\mathcal{P}(X_{i_1}, \dots, X_{i_n})) = (\mathcal{P}(X_{i_1+k}, \dots, X_{i_n+k})) .$$

For $d = q_{\min}$,

$$|H_m^{q_{\min}}| = |H_m^{\kappa+1}| = \begin{cases} m + 2^{\kappa-1}, & i \geq 0 \\ \frac{1}{2}(m + 2^{\kappa-1}) + \frac{3}{2}2^{\kappa}, & i < 0 \end{cases}$$

Theorem 5. *The size of the designs defined by (17) is*

$$\begin{aligned} |M_m^d| &= 1 + (c-1)|H_m^{q_{\min}}| + |H_m^t| \\ &= \begin{cases} \left(\left\lfloor \frac{d}{\kappa+1} \right\rfloor - 1\right)(m + 2^{\kappa-1}) + |H_m^t|, & i \geq 0 \\ \left(\left\lfloor \frac{d}{\kappa+1} \right\rfloor - 1\right)\left(\frac{1}{2}(m + 2^{\kappa-1}) + \frac{3}{2}2^{\kappa}\right) + |H_m^t|, & i < 0 \end{cases} \end{aligned}$$

where c , t and $q_{\min} = \kappa + 1$ are defined in eq. (16), and $|H_m^t|$ is given by Theorem 4.

Figure 4 shows the three families of graphs for $d = 19$ and $m = 5$, an example of a situation for which Morris' construction is not defined. Note the remarkably different graphs topologies, as well as the decreasing size: $76 = |G_5^{19}| > |H_5^{19}| = 60 > |M_5^{19}| = 49$.

We remark that for the values of d and m for which Morris designs are fully described in his paper, M_m^d is a perfectly equitable design of the same size, but our construction is defined for all pairs (d, m) .

3.4. Economy

Morris characterised efficiency of a design S as the ratio of the total number of elementary effects that can be computed using S to its size. We adhere to his definition.

Definition 3. Let $S \in \mathcal{E}_m^d$. The economy of S is

$$\Gamma(S) = \frac{md}{|S|}. \quad (18)$$

Lemma 8. $|G_m^d| \geq |H_m^d| \geq |M_m^d|$, $\Gamma(G_m^d) \leq \Gamma(H_m^d) \leq \Gamma(M_m^d)$.

Figure 5 confirms this Lemma: the economy of H_m^d is framed by the economies of G_m^d (below) and of M_m^d (above). These plots confirm that factorisation leads to a significant improvement, nearly doubling economy for small values of m .

Note (see right plot, where entire range of m for $d = 10$ is plotted) that all three curves come together at an economy of $d/2$ for the upper limit of $m = 2^{d-1}$, i.e., when the hypercube becomes complete. The point at which all curves merge is $m = 2^{d-2}$: for $m \geq 2^{d-2}$, (d, m) -edge equitable solutions are “unique,” in the sense that they all correspond to the deletion of $2^{d-1} - m$ non-adjacent points of Q_d . We can see that the middle curve (H_m^d) rapidly coincides with the (upper) curve for the factored design (M_m^d): since factored designs exist only if $d \geq 2q_{\min}$, i.e., for $m \leq 2^{(d/2-1)}$, from this point onwards the green and red curves are indistinguishable.

Lemma 9. When $d \rightarrow \infty$

$$\begin{aligned} \Gamma^\infty(G_m^d) &= \lim_{d \rightarrow \infty} \frac{md}{|G_m^d|} = 1 \\ \Gamma^\infty(H_m^d) &= \lim_{d \rightarrow \infty} \frac{md}{|H_m^d|} = \frac{m}{\alpha(m)} = \begin{cases} \frac{1}{1-2^{\kappa-1}/m}, & i \geq 0 \\ \frac{2}{1+2^{\kappa-1}/m}, & i < 0 \end{cases}, \quad \frac{4}{3} \leq \Gamma^\infty(H_m^d) \leq \frac{3}{2} \\ \Gamma^\infty(M_m^d) &= \lim_{d \rightarrow \infty} \frac{md}{|M_m^d|}, \quad \frac{m^2}{2m-1} \leq \Gamma^\infty(M_m^d) \leq \frac{m^2}{m-1} \end{aligned}$$

The expressions above follow from the definition of economy and the expressions for the size of the designs. Note that $\Gamma^\infty(G_m^d)$ is bounded below by $\Gamma^\infty(H_2^d) = 4/3$ and above by $\Gamma^\infty(H_3^d) = 3/2$. It is easy to check that $\Gamma^\infty(G_m^d) = \Gamma^\infty(G_1^d) = 1$, showing that the economy of our recursively defined solutions is bounded by the economy of their initialisations.

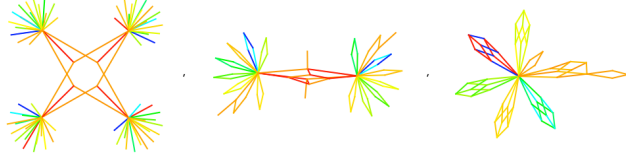


Figure 4: The three families of edge equitable graphs: G_5^{19} (left), H_5^{19} (centre) and M_5^{19} (right). Sizes are 76, 60 and 49, respectively.

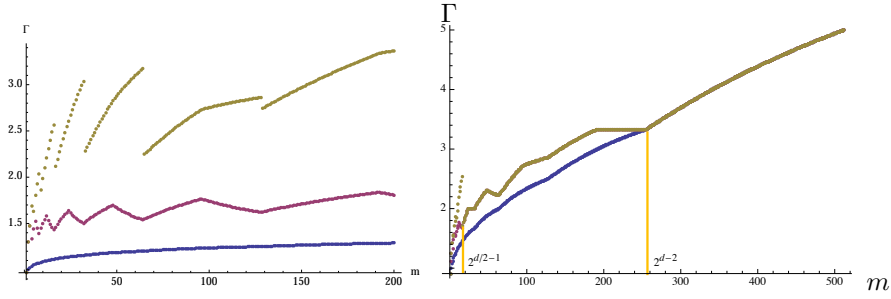


Figure 5: Economy (Γ) of designs G_m^d (lower curves), H_m^d (middle curves) and M_m^d (upper curves), $m \leq 200$, $d = 30$ (left) and complete range of m , $d = 10$ (right).

4. Sensitivity analysis

For One-At-a-Time designs, the elementary effects d_i can be computed incrementally as $f(\cdot)$ is evaluated at consecutive points of the design. This is no longer the case for $m > 1$. We indicate below how the polynomial representation can be exploited to identify the m pairs of points $\{(i_1^{(\ell)}, i_2^{(\ell)})\}_{\ell=1}^m$ involved in the computation of the m elementary effects along direction i . Let \mathcal{P} be the polynomial representation of the design and n be its size (the number of terms in \mathcal{P}). Consider an ordering of the monomials of \mathcal{P} , such that

$$\mathcal{P} = \sum_{p=1}^n \mathcal{P}_{s_p}, \quad s_p \in Q_d,$$

and let $f_{\mathcal{P}}$ denote the vector of valuations of $f(\cdot)$: $[f_{\mathcal{P}}]_p = f(s_p)$, $p = 1, \dots, n$. Define the d upper-triangular matrices

$$[E_i]_{(p,q)} = \begin{cases} (1 - 2[s_p]_i) \langle \mathcal{P}_{s_p}, X_i \mathcal{P}_{s_q} \rangle, & 1 \leq q < p \in \{1, \dots, n\}, i \in \{1, \dots, d\} \\ 0, & \text{otherwise} \end{cases}$$

There is at most one non-zero entry in each line of E . Assuming that \mathcal{P} is a (d, m) -edge equitable design, there are exactly m non-zero elements $[E_i]_{(i_1^{(\ell)}, i_2^{(\ell)})} = \pm 1, \ell = 1, \dots, m$, that indicate the pairs of points of the design that enable the computation of the m elementary effects d_i , which are the m non-zero entries of ($\mathbf{1}$ is the d -dimensional vector of 1's)

$$d_i = \frac{1}{\Delta} J_i f_{\mathcal{P}}, \quad J_i = -\text{diag}(E_i \mathbf{1}) + E_i .$$

Sample averages can be computed (remember there are only m non-zero values if d_i) as

$$\mu_i = \frac{1}{\Delta m} \mathbf{1}^T J_i f_{\mathcal{P}}, \quad \mu_i^* = \frac{1}{\Delta m} \mathbf{1}^T |J_i f_{\mathcal{P}}|, \quad i = 1, \dots, d .$$

Consider the following example in \mathcal{E}_2^4 :

$$\mathcal{P} = 1 + X_1 + X_2 + X_1 X_2 + X_1 X_2 X_3 + X_1 X_2 X_4 + X_1 X_2 X_3 X_4$$

For this graph, $n = 7$, and consider that the nodes are listed by order. Consider direction X_2 , for which the non-zero elements of J_2 are $[J_2]_{1,3} = [J_2]_{2,4} = 1$, and $[J_2]_{1,1} = [J_2]_{2,2} = -1$, and thus and

$$d_2^T = \frac{1}{\Delta} \begin{cases} 0, & i \notin \{1, 2\} \\ [f_{\mathcal{P}}]_3 - [f_{\mathcal{P}}]_1, & i = 4 \\ [f_{\mathcal{P}}]_4 - [f_{\mathcal{P}}]_2, & i = 6 \end{cases} .$$

Morris Elementary Effects method is based on a set of elementary effects computed along r random perturbations of a basic design \mathcal{P} . Using Lemma 6, random versions of a design represented by polynomial \mathcal{P} can be obtained as

$$\mathcal{P}^{(j)} = s^{(j)} \mathcal{P}(X^{(j)} + \pi^{(j)}(X)), \quad j = 1, \dots, r ,$$

where $\{s^{(j)}\}_{j=1}^r$ are independent and uniformly drawn in Q_d , $\{X^{(i)}\}_{j=1}^r$ are independent and uniform in \mathcal{A} and $\{\pi^{(j)}\}_{j=1}^r$ are independent random permutations of $\{1, \dots, d\}$.

5. Numerical application

We illustrate in this section the application of the designs presented in the previous sections, considering the same function as used in the original publication [10].

$$f(x) = \beta_0 + \sum_{i=1}^{20} \beta_i w_i + \sum_{i<j}^{20} \beta_{ij} w_i w_j + \sum_{i<j<l}^5 \beta_{ijl} w_i w_j w_l + \sum_{i<j<l<s}^4 \beta_{ijls} w_i w_j w_l w_s \quad (19)$$

where $w_i = 2X_i - 1, i \in \{1, 2, 4, 6, 8, \dots, 20\}$ and $w_i = 2.2X_i/(X_i + 0.1) - 1, i \in \{3, 5, 7\}$. Coefficients β_i are as follows:

$$\begin{aligned} \beta_i &= 20, & i &\in \{1, \dots, 10\}, & \beta_{ij} &= -15, & i, j &\in \{1, \dots, 6\} \\ \beta_{ijl} &= -10, & i, j, l &\in \{1, \dots, 5\}, & \beta_{ijls} &= 5, & i, j, l, s &\in \{1, \dots, 4\}. \end{aligned}$$

All remaining 1st and 2nd order coefficients are independent realisations of a standard normal distribution, $\beta_i \sim \mathcal{N}(0, 1), i \notin \{1, \dots, 10\}, \beta_{ij} \sim \mathcal{N}(0, 1), i, j \notin \{1, \dots, 6\}$. For this function the relevant classes of input factors are

$$\mathcal{C}_0 = \{11, \dots, 20\}, \quad \mathcal{C}_1 = \{8, 9, 10\}, \quad \mathcal{C}_2 = \{1, \dots, 7\}.$$

We apply Morris test for $m = 4$ and $r = 3$, leading to a total of number of elementary effects per direction equal to 12. The total number of evaluations of $f(\cdot)$ with this degree of multiplicity is $n_4 = 147$, while the computation of the same number of elementary effects with the standard designs ($m = 1$) requires $n_1 = 12(d + 1) = 252$, i.e., almost two times more.

Figure 6 shows the statistics observed in one run of the test. The estimates of the variances σ_i^2 have been corrected to take into account the clustered nature of designs (see [14] for details). The three distinct classes are well identified, although some class \mathcal{C}_2 input factors, like X_3 , come close to the \mathcal{C}_0 region. This tendency to wrongly classify non-linear/mixed effects, which can occasionally be classified as linear or negligible, has been recognised before, see [4]. In a subsequent paper we will fully address the study of Morris elementary method under clustered designs.

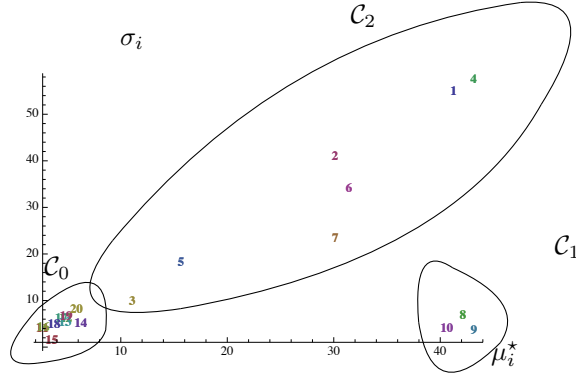


Figure 6: Morris statistics ($m = 4$) for $X_i, i = 1, \dots, 20$. Position of the label i indicates the observed (μ_i^*, σ_i) .

6. Conclusions and further work

The paper presents a complete and constructive definition of Morris clustered designs, that we designate by edge equitable designs. The algorithms presented are based on a polynomial representation of subgraphs of the hypercube that enables simple algebraic manipulation of the graphs and determination of their properties. These algorithms overcome some limitations of the original presentation: we provide recursive algorithms that enable the construction of equitable graphs with arbitrary multiplicity m for all dimensions d of the input space. The results are novel, and we are not aware of a formal study of this class of graphs in the literature.

Some extensions are possible. Our designs are not minimal, and the determination of minimal edge-equitable graphs remains an open problem. Our designs are subsets of the hypercube. The approach based on a polynomial representation of graphs may be extended to define equitable graphs over finite d -dimensional grids, recursively generated as the iterated product of a basic finite set S : $\mathcal{S}_d = \mathcal{S}_{d-1} \times S$, opening the way to computation of higher order derivatives. We are currently working in this direction.

Numerical studies show that there is a tradeoff between computational efficiency and discrimination power of the original Morris test when an increasing multiplicity m is used. However, use of $m > 1$ enables the definition of different kinds of tests, that will detect not just a large variability of the elementary effects across the entire domain

of f , but how much their distribution changes over disjoint neighborhoods of the input space. This idea will be explored in future studies.

Appendix A. Demonstration of Theorem 2

We consider separately the cases of odd and even m .

- m even

Assume that $G_{m/2}^{d-1} \in \mathcal{E}_{m/2}^{d-1}$, i.e.,

$$\langle G_{m/2}^{d-1}, X_i G_{m/2}^{d-1} \rangle = m, \quad i \in \{1, \dots, d-1\}.$$

Since $G_{m/2}^{d-1} \in K_{d-1}$,

$$\langle G_{m/2}^{d-1}, X_i X_d G_{m/2}^{d-1} \rangle = 0, \quad \forall i \neq d.$$

It follows immediately that, for G_m^d defined by (8)

$$\langle G_m^d, X_i G_m^d \rangle = \begin{cases} \langle G_{m/2}^{d-1}, X_i G_{m/2}^{d-1} \rangle + \\ \quad \langle X_1 X_d G_{m/2}^{d-1}, X_i X_1 X_d G_{m/2}^{d-1} \rangle, & \text{if } i < d \\ \langle G_{m/2}^{d-1}, X_1 G_{m/2}^{d-1} \rangle + \\ \quad \langle X_1 X_d G_{m/2}^{d-1}, X_1 G_{m/2}^{d-1} \rangle, & \text{if } i = d \end{cases}.$$

Each term in each branch of the right-hand-side of this equation is equal to m , demonstrating that $G_m^d \in \mathcal{E}_m^d$.

- m odd

Assume that $G_{m-1/2}^{d-1} \in \mathcal{E}_{m-1/2}^{d-1}$ and $G_{\frac{m+1}{2}}^{d-1}$. Then, for G_m^d defined by (9)

$$\begin{aligned} \langle G_m^d, X_i G_m^d \rangle &= \langle G_{m-1/2}^{d-1}, X_i G_{m-1/2}^{d-1} \rangle + \langle X_1 X_d G_{m+1/2}^{d-1}, X_i G_{m-1/2}^{d-1} \rangle \\ &\quad + \langle G_{m-1/2}^{d-1}, X_i X_1 X_d G_{m+1/2}^{d-1} \rangle + \langle X_1 X_d G_{m+1/2}^{d-1}, X_i X_1 X_d G_{m+1/2}^{d-1} \rangle \\ &= \begin{cases} \langle G_{m-1/2}^{d-1}, X_i G_{m-1/2}^{d-1} \rangle + \\ \quad + \langle G_{m+1/2}^{d-1}, X_i G_{m+1/2}^{d-1} \rangle, & \text{if } i < d \\ 2 \langle G_{m-1/2}^{d-1}, X_i X_1 X_d G_{m+1/2}^{d-1} \rangle, & \text{if } i = d \end{cases} \\ &= \begin{cases} (m-1) + (m+1) = 2m, & \text{if } i < d \\ 2 \langle G_{m-1/2}^{d-1}, X_1 G_{m+1/2}^{d-1} \rangle, & \text{if } i = d \end{cases} \end{aligned}$$

Construction (9) will thus lead to a (d, m) -edge equitable subgraph iff our family of solutions verifies also the following additional condition

$$\langle G_m^{d-1}, X_1 G_{m+1}^{d-1} \rangle = 2m + 1, \quad \forall m. \quad (.1)$$

Assume that $(m - 1)/2 = 2k$. Using equations (8) and (9),

$$\begin{aligned} \langle G_{2k}^d, X_1 G_{2k+1}^d \rangle &= \langle G_k^{d-1}, X_1 G_k^{d-1} \rangle + \langle G_k^{d-1}, X_d G_{k+1}^{d-1} \rangle \\ &\quad + \langle X_1 X_d G_k^{d-1}, X_1 G_k^{d-1} \rangle + \langle X_1 X_d G_k^{d-1}, X_d G_{k+1}^{d-1} \rangle \\ &= 2k + 0 + 0 + \langle X_1 G_k^{d-1}, G_{k+1}^{d-1} \rangle \end{aligned}$$

Equation (.1) will thus hold for $2k$ if it holds for k . When $(m - 1)/2 = 2k - 1$ we can easily check that the same implication is obtained:

$$\langle G_{k-1}^{d-1}, X_1 G_k^{d-1} \rangle = 2k - 1 \Rightarrow \langle G_{2k-1}^d, X_1 G_{2k}^d \rangle = 4k - 1$$

Thus, the condition for (9) to produce (d, m) -edge equitable solutions is

$$\forall d, \forall k \langle G_k^d, X_1 G_{k+1}^d \rangle = 2k + 1. \quad (.2)$$

It is easy to check that the condition holds for $k = 1$ (S_d) and $k = 2$ (the composition $(1 + X_1 X_d)S_d$), which concludes the proof.

Acknowledgment

This work has been partially funded by project Desire, ANR, *Programme Blanc International II* (FRANCE).

References

- [1] Boukouvalas, A., Gosling, J.P., Maruri-Aguilar, H., (2011). An Efficient Screening Method for Computer Experiments. *Tech. Report NCRG (Aston Univ.)*.
- [2] Campolongo, F., Braddock, R.D. (1999). The use of graph theory in the sensitivity analysis of the model output: a second-order screening method. *Reliability Eng. & System Safety* **64**, 1–12.

- [3] Campolongo, F., Cariboni, J., Saltelli, A. (2007). An effective screening design for sensitivity analysis of large models. *Environmental Modelling & Software* **22**, 1509–1518.
- [4] Campolongo, F., Saltelli, A., Cariboni, J. (2011). From screening to quantitative sensitivity analysis. A unified approach. *Computer Physics Communications* **182**, 978–988.
- [5] Cropp, R.A., Braddock, R.D., (2002). The New Morris method: an efficient second-order screening method. *Reliability Eng. & System Safety* **78**, 77–83.
- [6] Cukier, R.I., Fortuin, C.M., Shuler, K.E., Petschek, A.G., Schailby, H., (1973), Study of the Sensitivity of the Coupled Reaction Systems to Uncertainties in Rate Coefficients: I. Theory. *Journal of Chemical Physics*, **59**-8, 3873-3878.
- [7] Cukier, R.I., Levine, H.B., Schuler, K.E., (1978), Study of the Nonlinear sensitivity analysis of multiparameter model systems. *Journal of Chemical Physics*, **26**, 1-42.
- [8] Harary, F., Hayes, J. P., Wu, H.-J.(1988). A survey of subgraphs of the hypercube. *Comput. Math. Applic.* **15**-4, 277-289.
- [9] Horneberger, G.M., Spear, R.C., (1981). An approach to the preliminary analysis of environmental systems. *J. Env. Management*, **12**-7, 7-18.
- [10] Morris, M., (1991). Factorial Sampling Plans for Preliminary Computational Experiments. *Technometrics* **33**, 161–174.
- [11] Pistone, G., Wynn, H., (1996). Generalised confiding withGroßner Bases. *Biometrika* **83**-3, 653–666.
- [12] Pujol, (2007). Simplex-based screening designs for estimating metamodels. *Reliability Engineering and System Safety*, **94**-7, 1156-1160.
- [13] Saltelli, A., Ratto, M., Tarantola, S., Campolongo, F., (2006). Sensitivity Analysis Practices: Strategies for Model Based Inference. *Rel. Eng., & Syst. Safety* **91**, 1109–1125.

- [14] Skinner, C.J., (1981). Estimation of the Variance of a Finite Population for Cluster Samples. *Sankhyā: The Indian Journal of Statistics, Series B*, **43**-3, 392-398.
- [15] Sobol, I.M., (1993). Sensitivity estimates for nonlinear mathematical models. *Mathematical Modelling and Computational Experiments*, **1**-4, 407-414.
- [16] Williams, R.L., (1999). A Note on Robust Variance Estimation for Clustered-Correlated Data. *Biometrika* **56**, 645-646.